

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2006, 10:41:28 ; Search time 935 Seconds
(without alignments)
5100.554 Million cell updates/sec

Title: US-10-718-311-7

Perfect score: 684
Sequence: 1 atgggttcctctgtcatttc.....cggcgtaaccgttgactaa 684

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_8.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*
- 15: Geneseqn2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	100.0	684	6	ABA91838
2	684	100.0	684	6	ABA91838 Chloropla
3	684	100.0	684	10	Abk47712 DNA seque
4	684	100.0	684	10	Adc25988 Tomato/Es
5	501	73.2	3452	14	Adv91822 Tomato ru
6	499.4	73.0	1371	14	Adv91636 C. paraps
7	498	72.8	498	14	Adv91618 Escherich
8	498	72.8	854	4	Aad06956 ubiC gene
9	498	72.8	6641	4	Aad06957 pME2 comp
10	496.4	72.6	2000	2	Aat29820 E. coli u
11	495	72.4	495	2	Aaq92409 E. coli ch
12	495	72.4	495	6	ABA91837 Escherich
13	495	72.4	495	6	Abk47709 DNA seque
14	495	72.4	495	10	Adc25984 Escherich
15	472.8	69.1	1207	5	Aas90335 DNA encod
16	472.8	69.1	1318	2	Aaq67671 E.coli ub
17	282	41.2	645	11	ACH95439 Klebsiell
18	145.6	21.3	492	10	ADC76164 DNA homol

19	144	21.1	581	10	ADC75089	Adc75089 N bentham
20	144	21.1	614	10	ADC76956	Adc76956 DNA homol
21	144	21.1	632	10	ADC76953	Adc76953 DNA homol
22	144	21.1	718	10	ADK58382	Adk58382 Plant DNA
23	144	21.1	736	10	ADK54321	Adk54321 Plant DNA
24	144	21.1	736	10	ADK57660	Adk57660 Plant DNA
25	144	21.1	841	11	ADM44879	Adm44879 Insect re
26	144	21.1	847	11	ADM45447	Adm45447 Insect re
27	144	21.1	924	10	ADC75566	Adc75566 DNA homol
28	144	21.1	958	10	ADC76165	Adc76165 DNA homol
29	142.4	20.8	297	10	ADK59826	Adk59826 Plant DNA
30	142.4	20.8	377	10	ADK56121	Adk56121 Plant DNA
31	142.4	20.8	489	10	ADK56120	Adk56120 Plant DNA
32	142.4	20.8	504	10	ADK59824	Adk59824 Plant DNA
33	142.4	20.8	608	10	ADC76944	Adc76944 DNA homol
34	142.4	20.8	615	10	ADC76949	Adc76949 DNA homol
35	142.4	20.8	684	10	ADC76948	Adc76948 DNA homol
36	142.4	20.8	859	11	ADM45065	Adm45065 Insect re
37	141.4	20.7	619	11	ADM45158	Adm45158 Insect re
38	140.8	20.6	204	6	AAI70688	Aai70688 Tobacco R
39	140.8	20.6	499	6	AAD35635	Aad35635 Nicotiana
40	140.8	20.6	609	10	ADC76965	Adc76965 DNA homol
41	137.6	20.1	668	10	ADK59825	Adk59825 Plant DNA
42	137.6	20.1	668	11	ADM45687	Adm45687 Insect re
43	133.2	19.5	1442	2	AAQ92327	Aaq92327 Chloropla
44	132.6	19.4	171	14	ABE04209	Aeb04209 Tomato rb
45	132.6	19.4	171	14	ABE04212	Aeb04212 Tomato rb

ALIGNMENTS

RESULT 1

ABA91838

ID ABA91838 standard; DNA; 684 BP.

XX ABA91838;

XX 29-AUG-2003 (revised)

DT 15-MAY-2002 (first entry)

XX Chloroplast transit peptide-chorismate pyruvate lyase fusion gene.

XX Chloroplast transit peptide; tomato; Rubisco; plant;

KW ribulose-1,5-bisphosphate carboxylase; chorismate pyruvate lyase; CPL;

KW enzyme; p-hydroxybenzoic acid; transgenic plant; gene; ds.

XX Lycopersicon esculentum.

OS Escherichia coli.

OS Chimeric.

XX Key

FT CDS

Location/Qualifiers

1..495

/tag= a

/product= "chloroplast-targeted CPL fusion"

transit_peptide 1..186

/tag= b

/note= "tomato Rubisco chloroplast targeting sequence"

mat_peptide 187..492

/tag= c

/note= "E. coli chorismate pyruvate lyase"

WO200194607-A2.

13-DEC-2001.

22-MAY-2001; 2001WO-US016661.

02-JUN-2000; 2000US-0209854P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Meyer K, Van Dyk DE, Viitanen PV;

QY 421 ATTTGTTATGTGCGATGGTGAAACCGTGGCTTGGCGGTGCTGACCGTGGCTTCTGTGTCA 480
Db |||||||
QY 421 ATTTGTTATGTGCGGATGGTGAACCGTGGCTTGGCGGTGCTGACCGTGGCTTCTGTGTCA 480
Db |||||||
QY 481 ACGTTAAGCGGCGCGGAGCTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTAT 540
Db |||||||
QY 481 ACGTTAAGCGGCGCGGAGCTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTAT 540
Db |||||||
QY 541 CTGTTACATCATCGACATTAACCGGGACTTTATTGAGATAGGCGCTGATCCCGGGCTG 600
Db |||||||
QY 541 CTGTTACATCATCGACATTAACCGGGACTTTATTGAGATAGGCGCTGATCCCGGGCTG 600
Db |||||||
QY 601 TGGGGGCGACGTTCCCGCTCGATTAGCGGTAAACCGCTTGTGCTTAACAGAACTGTTT 660
Db |||||||
QY 601 TGGGGGCGACGTTCCCGCTCGATTAGCGGTAAACCGCTTGTGCTTAACAGAACTGTTT 660
Db |||||||
QY 661 TTACCGGCGTCAACCGTTGTACTAA 684
Db |||||||
QY 661 TTACCGGCGTCAACCGTTGTACTAA 684
Db |||||||

RESULT 4
ADV91622
ID ADV91622 standard; DNA; 684 BP.
XX
AC ADV91622;
XX
DT 10-MAR-2005 (first entry)
XX
DE Tomato rubisco small subunit precursor TP-E. coli CPL chimeric DNA.
XX
KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
KW genetically engineered microorganism; antioxidant; antimicrobial agent;
KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
KW neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid;
KW rubisco small subunit precursor; transit peptide;
KW chloroplast transit peptide; TP; chloroplast; gene; ds.
XX
OS Lycopersicon esculentum.
OS Escherichia coli.
OS Chimeric.
FH Key Location/Qualifiers
FT CDS 1..684
FT /*tag= a
FT /product= "Tomato rubisco small subunit precursor TP-E.
FT coli CPL fusion protein"

XX US2004261147-A1.
XX
XX 23-DEC-2004.
XX
XX 16-JUN-2003; 2003US-00462162.
XX
XX 16-JUN-2003; 2003US-00462162.
XX
XX (MEYE/) MEYER K.
XX PA (VIIT/) VIITANEN P V.
XX PA (FLIN/) FLINT D.
XX
XX Meyer K, Viitanen PV, Flint D;
XX
XX WPI; 2005-057232/06.
XX DR P-PSDB; ADV91623.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
XX green plant having nucleic acid fragments and chorismate pyruvate lyase
XX expression cassette.
XX
XX Example 4; SEQ ID NO 33; 70pp; English.

XX The present invention relates to methods and materials to produce
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
CC in genetically modified green plants and microorganisms. The method
CC relies upon transgenic plants or genetically modified microorganisms that
CC produce increased levels of the initial substrate para-hydroxybenzoic
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
CC arbutin. Plants and microbes can be genetically engineered to produce
CC high levels of pHBA either by functional expression of the bacterial
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
CC present sequence is tomato rubisco small subunit precursor (fbcS2)
CC transit peptide (chloroplast transit peptide; TP)-Escherichia coli CPL
CC chimeric DNA. This sequence is present in the E. coli expression vector
CC construct pET24a-TP-CPL.
XX
SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;
Query Match 100.0%; Score 684; DB 14; Length 684;
Best Local Similarity 100.0%; Pred. No. 1.4e-219;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGACCAATGTTACACAAGCT 60
Db |||||||
QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGACCAATGTTACACAAGCT 60
Db |||||||
QY 61 AGCATGGTTGCACCTTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db |||||||
QY 61 AGCATGGTTGCACCTTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db |||||||
QY 121 CAAACCTTGATCATCATCTCCATTGCTAGCAATGTTGGAAGTAGCTGCATGAGGTG 180
Db |||||||
QY 121 CAAACCTTGATCATCATCTCCATTGCTAGCAATGTTGGAAGTAGCTGCATGAGGTG 180
Db |||||||
QY 181 TGGCATATGTACACCCCGCGTTAACGCACTGCGTGGCGCTATTGTTAAAGAGATC 240
Db |||||||
QY 181 TGGCATATGTACACCCCGCGTTAACGCACTGCGTGGCGCTATTGTTAAAGAGATC 240
Db |||||||
QY 241 CTGCTCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300
Db |||||||
QY 241 CTGCTCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300
Db |||||||
QY 301 TTTGAAACAGCAGGGAACCGTAAAGCGTAGCATGATCCGGAAGGTTTGTTCGAGCAG 360
Db |||||||
QY 301 TTTGAAACAGCAGGGAACCGTAAAGCGTAGCATGATCCGGAAGGTTTGTTCGAGCAG 360
Db |||||||
QY 361 AATGAAATCCCGGAAGACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAA 420
Db |||||||
QY 361 AATGAAATCCCGGAAGACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAA 420
Db |||||||
QY 421 ATTTTGTATTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTGCTGCTGTGTCA 480
Db |||||||
QY 421 ATTTTGTATTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTGCTGCTGTGTCA 480
Db |||||||
QY 481 ACGTTAAGCGGCGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTAT 540
Db |||||||
QY 481 ACGTTAAGCGGCGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTAT 540
Db |||||||
QY 541 CTGTTACATCATCGACATTAACCGGGACTTTATTGAGATAGGCGCTGATCCCGGGCTG 600
Db |||||||
QY 541 CTGTTACATCATCGACATTAACCGGGACTTTATTGAGATAGGCGCTGATCCCGGGCTG 600
Db |||||||
QY 601 TGGGGGCGACGTTCCCGCTCGATTAGCGGTAAACCGCTTGTGCTTAACAGAACTGTTT 660
Db |||||||
QY 601 TGGGGGCGACGTTCCCGCTCGATTAGCGGTAAACCGCTTGTGCTTAACAGAACTGTTT 660
Db |||||||
QY 661 TTACCGGCGTCAACCGTTGTACTAA 684
Db |||||||
QY 661 TTACCGGCGTCAACCGTTGTACTAA 684
Db |||||||

```

RESULT 5
ADV91636
XX ADV91636 standard; DNA; 3452 BP.
XX AC
XX ADV91636;
XX DT 10-MAR-2005 (first entry)
XX C. parapsilosis pHB 1-H- E. coli CPL- A. thaliana UGT72B1 chimeric DNA.
XX DE
XX KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
XX KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
XX KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
XX KW genetically engineered microorganism; antioxidant; antimicrobial agent;
XX KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
XX KW neoplasm; pHBA 1-hydroxylase; pHBA 1-H; pHBA; para-hydroxybenzoic acid;
XX KW p-hydroxybenzoic acid; UDP-glucosyltransferase; UGT72B1; gene; ds.
XX OS Arabidopsis thaliana.
XX OS Candida parapsilosis.
XX OS Escherichia coli.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX FT CDS 4..501
XX FT /*tag= a
XX FT /product= "Escherichia coli CPL protein"
XX FT 541..1980
XX FT /*tag= b
XX FT /product= "Candida parapsilosis pHBA 1-H protein"
XX FT /transl_except= (pos: 559..561, aa:Trp)
XX FT CDS 2004..3446
XX FT /*tag= c
XX FT /product= "Arabidopsis thaliana UGT72B1 protein"
XX US2004261147-A1.
XX PN
XX PD 23-DEC-2004.
XX XX
XX PF 16-JUN-2003; 2003US-00462162.
XX XX
XX PR 16-JUN-2003; 2003US-00462162.
XX XX
XX PA (MEYE/) MEYER K.
XX PA (VIIT/) VIITANEN P V.
XX PA (FLIN/) FLINT D.
XX XX
XX PI Meyer K, Viitanen PV, Flint D;
XX XX
XX DR WPI; 2005-057232/06.
XX DR P-PSDB; ADV91605, ADV91619, ADV91631.
XX XX
XX PT Producing hydroquinone glucoside in a green plant comprises growing a
XX PT green plant having nucleic acid fragments and chorismate pyruvate lyase
XX PT expression cassette.
XX XX
XX PS Example 7; SEQ ID NO 47; 70pp; English.
XX XX
XX CC The present invention relates to methods and materials to produce
XX CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX CC in genetically modified green plants and microorganisms. The method
XX CC relies upon transgenic plants or genetically modified microorganisms that
XX CC produce increased levels of the initial substrate para-hydroxybenzoic
XX CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
XX CC arbutin. Plants and microbes can be genetically engineered to produce
XX CC high levels of pHBA either by functional expression of the bacterial
XX CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
XX CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
XX CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
XX CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
XX CC present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) -
XX CC Escherichia coli chorismate pyruvate lyase (CPL) and Arabidopsis thaliana
XX CC UDP-glucosyltransferase (UGT72B1) chimeric DNA.
XX CC

XX SQ Sequence 3452 BP; 932 A; 730 C; 856 G; 934 T; 0 U; 0 Other;
XX Query Match 73.2%; Score 501; DB 14; Length 3452;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-157;
XX Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX DT 10-MAR-2005 (first entry)
XX C. parapsilosis pHB 1-H- E. coli CPL- A. thaliana UGT72B1 chimeric DNA.
XX DE
XX KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
XX KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
XX KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
XX KW genetically engineered microorganism; antioxidant; antimicrobial agent;
XX KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
XX KW neoplasm; pHBA 1-hydroxylase; pHBA 1-H; pHBA; para-hydroxybenzoic acid;
XX KW p-hydroxybenzoic acid; UDP-glucosyltransferase; UGT72B1; gene; ds.
XX OS Arabidopsis thaliana.
XX OS Candida parapsilosis.
XX OS Escherichia coli.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX FT CDS 4..501
XX FT /*tag= a
XX FT /product= "Escherichia coli CPL protein"
XX FT 541..1980
XX FT /*tag= b
XX FT /product= "Candida parapsilosis pHBA 1-H protein"
XX FT /transl_except= (pos: 559..561, aa:Trp)
XX FT CDS 2004..3446
XX FT /*tag= c
XX FT /product= "Arabidopsis thaliana UGT72B1 protein"
XX US2004261147-A1.
XX PN
XX PD 23-DEC-2004.
XX XX
XX PF 16-JUN-2003; 2003US-00462162.
XX XX
XX PR 16-JUN-2003; 2003US-00462162.
XX XX
XX PA (MEYE/) MEYER K.
XX PA (VIIT/) VIITANEN P V.
XX PA (FLIN/) FLINT D.
XX XX
XX PI Meyer K, Viitanen PV, Flint D;
XX XX
XX DR WPI; 2005-057232/06.
XX DR P-PSDB; ADV91605, ADV91619, ADV91631.
XX XX
XX PT Producing hydroquinone glucoside in a green plant comprises growing a
XX PT green plant having nucleic acid fragments and chorismate pyruvate lyase
XX PT expression cassette.
XX XX
XX PS Example 7; SEQ ID NO 47; 70pp; English.
XX XX
XX CC The present invention relates to methods and materials to produce
XX CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX CC in genetically modified green plants and microorganisms. The method
XX CC relies upon transgenic plants or genetically modified microorganisms that
XX CC produce increased levels of the initial substrate para-hydroxybenzoic
XX CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
XX CC arbutin. Plants and microbes can be genetically engineered to produce
XX CC high levels of pHBA either by functional expression of the bacterial
XX CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
XX CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
XX CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
XX CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
XX CC present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) -
XX CC Escherichia coli chorismate pyruvate lyase (CPL) and Arabidopsis thaliana
XX CC UDP-glucosyltransferase (UGT72B1) chimeric DNA.
XX CC

```

FT /product= "Candida parapsilosis pHBA 1-H protein"
FT /transl_except= (pos: 540..542, aa:Trp)
XX
PN US2004261147-A1.
XX
PD 23-DEC-2004.
XX
XX 16-JUN-2003; 2003US-00462162.
PF 16-JUN-2003; 2003US-00462162.
PR
PA (MEYE/) MEYER K.
PA (VIIT/) VIITANEN P V.
PA (FLIN/) FLINT D.
XX
PI Meyer K, Viitanen PV, Flint D;
XX WPI; 2005-057232/06.
DR P-PSDB; ADV91605, ADV91619.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
PT green plant having nucleic acid fragments and chorismate pyruvate lyase
PT expression cassette.
XX
PS Example 7; SEQ ID NO 54; 70pp; English.
XX
CC The present invention relates to methods and materials to produce
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
CC in genetically modified green plants and microorganisms. The method
CC relies upon transgenic plants or genetically modified microorganisms that
CC produce increased levels of the initial substrate para-hydroxybenzoic
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
CC arbutin. Plants and microbes can be genetically engineered to produce
CC high levels of pHBA either by functional expression of the bacterial
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
CC present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) and
CC Escherichia coli chorismate pyruvate lyase (CPL) chimeric DNA.
XX
SQ Sequence 1971 BP; 550 A; 383 C; 484 G; 554 T; 0 U; 0 Other;

Query Match 73.0%; Score 499.4; DB 14; Length 1971;
Best Local Similarity 99.8%; Pred. No. 4.9e-157;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 184 CATATGTACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 243
DB 1 CATATGTACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 60

QY 244 GCCCTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTT 303
DB 61 GCCCTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTT 120

QY 304 GAACAGCAGGGAACCGTAAGCTGACGATGATCCGGAAGGTTTGTGAGCAGGAT 363
DB 121 GAACAGCAGGGAACCGTAAGCTGACGATGATCCGGAAGGTTTGTGAGCAGGAT 180

QY 364 GAAATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTAAT 423
DB 181 GAAATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTAAT 240

QY 424 TTGTTATGTGCGGATGGTGAACCGTGGCTGTGCGGTGCTGACGTCGTTCCCTGTGTCAACG 483
DB 241 TTGTTATGTGCGGATGGTGAACCGTGGCTGTGCGGTGCTGACGTCGTTCCCTGTGTCAACG 300

QY 484 TTAACCGGCGGAGCTCGGCTTACAAAATTTGGGTAAACGCGTTAGGACGCTATCTG 543
DB 301 TTAACCGGCGGAGCTCGGCTTACAAAATTTGGGTAAACGCGTTAGGACGCTATCTG 360

QY 544 TTCACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGCGTGTGTCGGGCTGTGG 603
|||||

DB 361 TTCACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGCGTGTGCTGTGG 420
QY 604 GGGGACGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTA 663
XX |||||||
DB 421 GGGGACGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTA 480
XX |||||||
QY 664 CCGGCGTCACCGTTGTACTAA 684
DB 481 CCGGCGTCACCGTTGTACTAA 501
XX |||||||

RESULT 7
ADV91618
ID ADV91618 standard; DNA; 498 BP.
XX AC
XX ADV91618;
XX AC
XX 10-MAR-2005 (first entry)
XX DE
XX Escherichia coli chorismate pyruvate lyase (CPL) DNA.
XX
XX Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
KW genetically engineered microorganism; antioxidant; antimicrobial agent;
KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
KW neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid; gene;
KW ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..498
FT /*tag= a
FT /product= "Escherichia coli CPL protein"
XX
XX US2004261147-A1.
PD 23-DEC-2004.
XX
XX 16-JUN-2003; 2003US-00462162.
PF
XX 16-JUN-2003; 2003US-00462162.
XX
XX (MEYE/) MEYER K.
XX (VIIT/) VIITANEN P V.
XX (FLIN/) FLINT D.
XX
XX Meyer K, Viitanen PV, Flint D;
XX WPI; 2005-057232/06.
DR P-PSDB; ADV91619.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
PT green plant having nucleic acid fragments and chorismate pyruvate lyase
PT expression cassette.
XX
XX Example 4; SEQ ID NO 29; 70pp; English.
XX
XX The present invention relates to methods and materials to produce
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
CC in genetically modified green plants and microorganisms. The method
CC relies upon transgenic plants or genetically modified microorganisms that
CC produce increased levels of the initial substrate para-hydroxybenzoic
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
CC arbutin. Plants and microbes can be genetically engineered to produce
CC high levels of pHBA either by functional expression of the bacterial
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
CC present sequence is Escherichia coli chorismate pyruvate lyase (CPL) DNA.
CC This sequence is found in the E. coli expression vector construct

[illegible]

RESULT 8	
AAAD06956	
ID	AAAD06956 standard; DNA; 854 BP.
XX	
XX	
AC	AAAD06956;
XX	
XX	
DT	06-AUG-2001 (first entry)
XX	
DE	ubiC gene encoding chorismate pyruvate lyase.
XX	
KW	p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;
KW	chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;
KW	ppp synthase; tktA; transketolase; aroC; chorismate synthase; aroA;
KW	5-enolpyruvylshikimate-3-phosphate synthase; EPSP synthase; aroL;
KW	shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;
KW	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHQ synthase; ds.
XX	
OS	Escherichia coli.
XX	
PN	US6210937-B1.
XX	
PD	03-APR-2001.
XX	
PF	22-APR-1998; 98US-00064693.
XX	
PR	22-APR-1997; 97US-0044094P.
XX	
PA	(BECH-) BECHTEL BWXT IDAHO LLC.

XX	Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;
PI	WPI; 2001-280857/29.
XX	
DR	
XX	
PT	Producing aromatic compounds, especially commercially acceptable levels
PT	of p-hydroxybenzoic acid, comprises developing genetically engineered
PT	bacteria that carry selected genes of the common aromatic pathway.
XX	
PS	Claim 5; Col 31-32; 25pp; English.
PS	
XX	
CC	The present invention relates to a method of producing p-hydroxybenzoic
CC	acid (PHB) which comprises transforming microorganisms with plasmids
CC	carrying selected genes of the common aromatic pathway and by conversion
CC	of carbon sources. The selected genes of the common aromatic pathway of
CC	the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG
CC	structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate
CC	(DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppvA
CC	gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding,
CC	chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene
CC	encoding 5-enolpyruvoylshikimate-3-phosphate (EPSP) synthase and aroB
CC	gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for
CC	converting chorismate to a selected aromatic compound. The method is
CC	particularly used for the production of commercially acceptable levels of
CC	aromatic compounds, particularly p-hydroxybenzoic acid. The present
CC	sequence is ubiC gene which encodes chorismate pyruvate lyase that
CC	catalyses the production of chorismate from PHB
XX	
SQ	Sequence 854 BP; 202 A; 200 C; 222 G; 230 T; 0 U; 0 Other;
	Query Match 72.8%; Score 498; DB 4; Length 854;
	Best Local Similarity 100.0%; Pred. No. 9,4e-157;
	Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	187 ATGTCAACCCCSCGGTTAACGCAACTCGCGTCGCTATTGTAAGAGATCCCTGCC 246
Db	
	321 ATGTCAACCCCSCGGTTAACGCAACTCGCGTCGCTATTGTAAGAGATCCCTGCC 380
Qy	247 CTGGATCCGCAACTGCTCGACTGGCTTGTCTCGAGGATTCCATGACAAAACGTTTTGAA 306
Db	
	381 CTGGATCCGCAACTGCTCGACTGGCTTGTCTCGAGGATTCCATGACAAAACGTTTTGAA 440
Qy	307 CAGCAGGGAAAAACGGTAGCGTGACGATGATCCGGAAGGGTTGTGCGAGCAGAATGAA 366
Db	
	441 CAGCAGGGAAAAACGGTAGCGTGACGATGATCCGGAAGGGTTGTGCGAGCAGAATGAA 500
Qy	367 ATCCCGAAGAACTGGCGTGTCTGCCGAAAGAGTCTGTTACTGGTTACGTGAATTTTG 426
Db	
	501 ATCCCGAAGAACTGGCGTGTCTGCCGAAAGAGTCTGTTACTGGTTACGTGAATTTTG 560
Qy	427 TTATGTGCCGATGGTGAAACCGTGGCTTGC CGGTGCTACCGTCTCTCTGTGTCAAACGTTA 486
Db	
	561 TTATGTGCCGATGGTGAAACCGTGGCTTGC CGGTGCTACCGTCTCTCTGTGTCAAACGTTA 620
Qy	487 AGCGGGCCGAGCTGGCGTTACAAAAATTGGGTAAAAACGCCGTATAGGACGCTATCTGTT 546
Db	
	621 AGCGGGCCGAGCTGGCGTTACAAAAATTGGGTAAAAACGCCGTATAGGACGCTATCTGTT 680
Qy	547 ACATCATCGACATTACC CGGGACTTTATTGAGATAGCGCGTATCGCGGCTGTCGGGG 606
Db	
	681 ACATCATCGACATTACC CGGGACTTTATTGAGATAGCGCGTATCGCGGCTGTCGGGG 740
Qy	607 CGACGTTTCCCGCTCGGATTAAAGCGGTAAACCCCTGTTGCTAAACAGAACTGTTTTTACCG 666
Db	
	741 CGACGTTTCCCGCTCGGATTAAAGCGGTAAACCCCTGTTGCTAAACAGAACTGTTTTTACCG 800
Qy	667 GCCTCACCGTTGACTAA 684
Db	
	801 GCCTCACCGTTGACTAA 818

RESULT 9
AAD06957

XX The present sequence comprises the ubiquinone biosynthetic enzyme genes
 CC ubiC and ubiA, which were cloned from the chromosomal DNA of *E. coli*
 CC (Kohara map phage DNA bank IF8 (634); Cell 50, 495-508 (1987)). A large
 CC amt. of ubiquinone-10 can be produced by culturing photosynthetic
 CC bacteria transformed with the novel plasmid pRSFAC, which contains the
 CC ubiC and ubiA genes and the glutamate synthase gene promoter,
 CC specifically the Rhodospirillum rubrum transformant MC9R/pRSFAC
 XX
 SQ Sequence 2000 BP; 444 A; 462 C; 538 G; 556 T; 0 U; 0 Other;

Query Match 72.6%; Score 496.4; DB 2; Length 2000;
 Best Local Similarity 99.8%; Pred. No. 5.1e-156;
 Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 187 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCTGCTATTGTAAGAGATCCCTGCC 246
 DB 380 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCTGCTATTGTAAGAGATCCCTGCC 439
 QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 306
 DB 440 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 499
 QY 307 CAGCAGGGAACCGTACGATGATCCGGAAGGGTTTGTGAGCAGAGATGAA 366
 DB 500 CAGCAGGGAACCGTACGATGATCCGGAAGGGTTTGTGAGCAGAGATGAA 559
 QY 367 ATCCCGGAAGACTGCGCTGCTGCGGAAGAGTCTGTTACTGTTAGTGAATTTG 426
 DB 560 ATCCCGGAAGACTGCGCTGCTGCGGAAGAGTCTGTTACTGTTAGTGAATTTG 619
 QY 427 TTATGTGCGGATGGTGAACCGTGGCTTGGCGTACCGTCTGTTCTGTGCAACGTTA 486
 DB 620 TTATGTGCGGATGGTGAACCGTGGCTTGGCGTACCGTCTGTTCTGTGCAACGTTA 679
 QY 487 AGCGGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGGACGTTATCTGTTTC 546
 DB 680 AGCGGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGGACGTTATCTGTTTC 739
 QY 547 ACATCATCAGATTAACCGGAGCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 606
 DB 740 ACATCATCAGATTAACCGGAGCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 799
 QY 607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 666
 DB 800 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 859
 QY 667 GCGTACCGTTGTTACTAA 684
 DB 860 GCGTACCGTTGTTACTAA 877

RESULT 11

AAQ92409
 ID AAQ92409 standard; DNA; 495 BP.

XX
 AC
 XX
 XX
 XX

DT 11-JAN-1996 (first entry)

XX
 DE *E. coli* chorismate-pyruvate lyase coding sequence.

KW Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;
 KW tobacco mosaic virus resistance; antiviral activity;
 KW p-hydroxybenzoic acid; ss.

XX
 OS *Escherichia coli*.

XX
 PN DE4423022-Cl.

XX
 PD 24-MAY-1995.

XX
 PF 30-JUN-1994; 94DE-04423022.

XX 30-JUN-1994; 94DE-04423022.
 XX (HEID/) HEIDE L.
 XX Heide L, Siebert M, Severin K;
 XX WPI; 1995-186908/25.
 XX P-PSDB; AAR74742.
 XX Transgenic plants with increased content of resistance factor - contg.
 PT e.g. antiviral p-hydroxy-benzoic acid.
 XX
 PS Claim 3; Page 5; 7pp; German.

XX Transgenic plants containing a bacterial gene which encodes an enzyme
 CC able to catalyse prodn. of an antiviral, bactericidal, fungicidal or
 CC insecticidal factor are new. A preferred gene is ubiC from *E. coli*
 CC (AAQ92409) which codes for chorismate-pyruvate lyase (AAR74742). The
 CC lyase catalyses conversion of chorismate to p-hydroxybenzoic acid and
 CC transgenic tobacco plants which express the active enzyme are resistant
 CC to tobacco mosaic virus

XX SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Query Match 72.4%; Score 495; DB 2; Length 495;
 Best Local Similarity 100.0%; Pred. No. 7.2e-156;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCTGCTATTGTAAGAGATCCCTGCC 246
 DB 1 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCTGCTATTGTAAGAGATCCCTGCC 60
 QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 306
 DB 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 120
 QY 307 CAGCAGGGAACCGTACGATGATCCGGAAGGGTTTGTGAGCAGAGATGAA 366
 DB 121 CAGCAGGGAACCGTACGATGATCCGGAAGGGTTTGTGAGCAGAGATGAA 180
 QY 367 ATCCCGGAAGACTGCGCTGCTGCGGAAGAGTCTGTTACTGTTAGTGAATTTG 426
 DB 181 ATCCCGGAAGACTGCGCTGCTGCGGAAGAGTCTGTTACTGTTAGTGAATTTG 240
 QY 427 TTATGTGCGGATGGTGAACCGTGGCTTGGCGTACCGTCTGTTCTGTGCAACGTTA 486
 DB 241 TTATGTGCGGATGGTGAACCGTGGCTTGGCGTACCGTCTGTTCTGTGCAACGTTA 300
 QY 487 AGCGGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 546
 DB 301 AGCGGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
 QY 547 ACATCATCAGATTAACCGGAGCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 606
 DB 361 ACATCATCAGATTAACCGGAGCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 420
 QY 607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 666
 DB 421 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 480
 QY 667 GCGTACCGTTGTTACTAC 681
 DB 481 GCGTACCGTTGTTACTAC 495

RESULT 12

ABA91837
 ID ABA91837 standard; DNA; 495 BP.

XX
 AC ABA91837;
 XX

```

DT 15-MAY-2002 (first entry)
XX Escherichia coli chorismate pyruvate lyase gene coding region.
DE
XX Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;
KW transgenic plant; ubiC; gene; ds.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH CDS 1..495
FT /*tag= a
FT /partial
FT /product= "Chorismate_pyruvate_lyase"
FT /note= "the CDS does not include a stop codon"
XX
XX WO200194607-A2.
XX
XX PD 13-DEC-2001.
XX
XX PF 22-MAY-2001; 2001WO-US016661.
XX
XX PR 02-JUN-2000; 2000US-0209854P.
XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Meyer K, Van Dyk DE, Viitanen PV;
XX
XX DR WPI; 2002-226795/28.
XX
XX DR P-PSDB; AAM50958.
XX
XX PT Producing para-hydroxy benzoic acid in green plant, comprises expression
XX of unique expression cassette containing gene encoding chorismate
XX pyruvate lyase operably linked to specific chloroplast targeting
XX sequence.
XX
XX PS Claim 4; Page 55; 60pp; English.
XX
XX CC The present sequence is that of the coding region of the Escherichia coli
XX strain W3110 ubiC gene in expression construct pET24a. The gene encodes
XX chorismate pyruvate lyase (CPL, see AAM50968). It was obtained by PCR
XX amplification of strain W3110 genomic DNA using primers (see ABA91839-40)
XX based on the published E. coli ubiC gene. The CPL open reading frame is
XX used in an expression cassette designed for the high-level production of
XX p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette
XX comprises the CPL coding sequence operably linked to a promoter capable
XX of driving protein expression in higher plants. The cassette also has a
XX sequence encoding a chloroplast transit peptide, its natural cleavage
XX site, and a small portion of a transit peptide donor protein fused to the
XX N-terminus of CPL. The chloroplast targeting sequence targets the foreign
XX protein to the chloroplast and aids in its uptake into the organelle. The
XX cleavage site is unique to the transit peptide, and cleavage of the
XX chimeric protein encoded by the cassette at this site releases a novel
XX polypeptide that has full enzyme activity, comprising the mature CPL
XX enzyme and a small portion of transit peptide donor (see AAM50961). A
XX plant comprising the CPL expression cassette is claimed, and may be
XX soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
XX barley, oats, sorghum, rice, Arabidopsis, sugarbeet, sugarcane, canola,
XX millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
XX component of liquid crystal polymers which have application in the
XX automotive, electrical and other industries
XX
XX SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;
XX
XX Query Match 72.4%; Score 495; DB 6; Length 495;
XX Best Local Similarity 100.0%; Pred. No. 7.2e-156;
XX Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 187 ATGTCACACCCCGCGTTAACGCAACTCGTGGCTGCTGCTGAGGAGTTCATGACAAACGTTTGA 246
XX 1 ATGTCACACCCCGCGTTAACGCAACTCGTGGCTGCTGCTGAGGAGTTCATGAGAGATCCCTGCC 60
XX
XX QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAACGTTTGA 306

```

```

Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 120
QY 307 CAGCAGGGAACAAACGGTAAAGCGTGACGATGATCCGCGAAGGTTTGTCCGAGCAGATGAA 366
Db 121 CAGCAGGGAACAAACGGTAAAGCGTGACGATGATCCGCGAAGGTTTGTCCGAGCAGATGAA 180
QY 367 ATCCCGGGAAGAACTGCGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426
Db 181 ATCCCGGGAAGAACTGCGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240
QY 427 TTATGTGCCGATGTTGAACCGTGGCTTCCGCGTACCGTCTTCTGTTGTCACGTTA 486
Db 241 TTATGTGCCGATGTTGAACCGTGGCTTCCGCGTACCGTCTTCTGTTGTCACGTTA 300
QY 487 AGCGGGCGGAGCTGGCGTTTACAAAATTGGGTTAAACCGCGTTAGGACGCTATCTGTTTC 546
Db 301 AGCGGGCGGAGCTGGCGTTTACAAAATTGGGTTAAACCGCGTTAGGACGCTATCTGTTTC 360
QY 547 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGCCCGTGTATGCCGGCTGTGGGG 606
Db 361 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGCCCGTGTATGCCGGCTGTGGGG 420
QY 607 CGAGCTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTAAACAGAACTGTTTTCACCG 666
Db 421 CGAGCTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTAAACAGAACTGTTTTCACCG 480
QY 667 GCGTCACCGTTGTATC 681
Db 481 GCGTCACCGTTGTATC 495

RESULT 13
ABK47709
ID ABK47709 standard; DNA; 495 BP.
XX
XX AC ABK47709;
XX
XX DT 18-JUN-2002 (first entry)
XX
XX DE DNA sequence of ORF for CPL in pET24a E. coli expression construct.
XX
XX KW SNGI; sinapoylglucose accumulator 1; SMT; aromatic ester;
XX sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;
XX malate conjugated aromatic acid; polymer synthesis; ubiC; CPL; gene;
XX carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;
XX pET24a; ds.
XX
XX OS Escherichia coli.
XX
XX FH Key Location/Qualifiers
XX CDS 1..495
XX FT /*tag= a
XX FT /partial
XX FT /product= "CPL"
XX FT /note= "This sequence lacks a stop codon"
XX
XX WO200204653-A2.
XX
XX PD 17-JAN-2002.
XX
XX PF 05-JUL-2001; 2001WO-US021283.
XX
XX PR 07-JUL-2000; 2000US-0216615P.
XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Flint D, Meyer K, Viitanen PV;
XX
XX DR WPI; 2002-303779/34.
XX
XX DR P-PSDB; AAU77941.
XX
XX PT Producing aromatic acid conjugates, involves contacting glycosylated

```

.....

KW UDP-glucosyltransferase; pHBA; p-hydroxybenzoic acid ester glucoside;
 KW liquid crystal polymer; LCP; methylparaben; preservative; food;
 KW cosmetic industry; ds; gene; ubic; CPL; chorismate pyruvate lyase.

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: May 29, 2006, 20:59:50 ; Search time 746.985 Seconds
(without alignments)
2380.132 Million cell updates/sec

Title: US-10-718-311-16
Perfect score: 887
Sequence: 1 MQVWMSHPALQRLALRYC.....RLSGKPLLLTELFLPASPLY 170

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5244920 seqs, 3486124231 residues
Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-Q=/abs/ABSSWEB spool/US10718311/runat_26052006_164914_13428/app.query.fasta_1
-DB=N_Geneseq -QFM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.0.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US10718311 @CN 1_1 2295 @runat_26052006_164914_13428 -NCFU=6 -ICPU=3
-NO WMAP -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.8.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*
- 15: Geneseqn2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	887	100.0	684	6 ABA91838	Chloroplast
2	887	100.0	684	6 ABK47712	Abk47712 DNA sequ
3	887	100.0	684	10 ADC25988	Adc25988 Tomato/Es

4	887	100.0	684	14 ADV91622	Adv91622 Tomato ru
5	882	97.2	3452	14 ADV91636	Adv91636 C. paraps
6	854	96.3	495	2 AAQ92409	Aaq92409 E.coli ch
7	854	96.3	495	6 ABA91837	Abk47709 DNA sequ
8	854	96.3	495	6 ABK47709	Abk47709 DNA sequ
9	854	96.3	495	10 ADC25984	Adc25984 Escherich
10	854	96.3	498	14 ADV91618	Adv91618 Escherich
11	854	96.3	854	4 AAD06956	Aad06956 ubiC gene
12	854	96.3	6641	4 AAD06957	Aad06957 pHE2 comp
13	851	95.9	1971	14 ADV91643	Adv91643 C. paraps
14	851	95.9	2000	2 AAT29820	Aat29820 E. coli u
15	811	91.4	1207	5 AAS90335	Aas90335 DNA encod
16	791	89.2	1918	2 AAQ67671	Aaq67671 E.coli ub
17	673	75.9	645	11 ACH95439	Ach95439 Klebsiell
18	417	47.0	519	10 ACF70957	Acf70957 Photorhab
19	417	47.0	110000	10 ACF67367_40	Continuation (41 o
20	417	47.0	110000	10 ACF65388_07	Continuation (8 of
21	401	45.2	537	10 ADF02837	Adf02837 Bacterial
22	186	21.0	630	11 ABD08838	Abd08838 Pseudomon
23	174.5	19.7	519	12 ADL04203	Adl04203 DNA encod
24	174	19.6	100	8 ACD78861	AcD78861 E. coli K
25	171.5	19.3	269223	4 AAF28554	Aaf28554 Genomic f
26	167	18.8	100	8 ACD78860	AcD78860 E. coli K
27	165	18.6	100	8 ACD78859	AcD78859 E. coli K
28	159	17.9	540	9 ADA29056	Ada29056 DNA encod
29	146.5	16.5	1248	11 ABD08757	Abd08757 Pseudomon
30	133.5	15.1	55839	12 ADN36893	Adn36893 X. albili
31	130.5	14.7	609	12 ADN38906	Adn38906 X. albili
32	121	13.6	195	10 ADF02801	Adf02801 Bacterial
33	97.5	11.0	475	6 AAS97252	Aas97252 Neisseria
34	90	10.1	2127	4 ABA95459	Aba95459 Thermus t
35	90	10.1	9133	3 AAA81730	Aaa81730 N. mening
36	90	10.1	110000	3 AAA81490_07	Continuation (8 of
37	90	10.1	349880	3 AAF21608	Aaf21608 Neisseria
38	88.5	10.0	2892	15 AEF63232	Aef63232 Human UNC
39	88	9.9	42811	12 ADQ97948	Adq97948 Human can
40	87	9.8	59589	11 ACN44724	Acn44724 Mouse gen
41	86.5	9.8	110000	10 ACF67367_43	Continuation (44 o
42	86.5	9.8	110000	10 ACF65388_04	Continuation (5 of
43	86	9.7	522	10 ABZ33133	Abz33133 N. gonorr
44	85.5	9.6	2157	1 AAN70438	Aan70438 Streptoco
45	85.5	9.6	2907	11 ADM02382	Adm02382 Human CDN

ALIGNMENTS

RESULT 1
ID ABA91838 standard; DNA; 684 BP.
XX ABA91838;
AC ABA91838;

DT 29-AUG-2003 (revised)
DT 15-MAY-2002 (first entry)
XX

Chloroplast transit peptide-chorismate pyruvate lyase fusion gene.

Chloroplast transit peptide; tomato; Rubisco; plant;

ribulose-1,5-bisphosphate carboxylase; chorismate pyruvate lyase; CPL;
enzyme; p-hydroxybenzoic acid; transgenic plant; gene; ds.

Lycopersicon esculentum.

Escherichia coli.

Chimeric.

Key Location/Qualifiers
CDS 1..495
/tag= a
/product= "chloroplast-targeted CPL fusion"

transit_peptide 1..186
/tag= b

mat_peptide 187..492
/note= "tomato Rubisco chloroplast targeting sequence"

```

FT      /*tag= C
FT      /note= "E. coli chorismate pyruvate lyase"
PN      WO200194607-A2.
XX      13-DEC-2001.
XX
XX      22-MAY-2001; 2001WO-US016661.
XX
XX      02-JUN-2000; 2000US-0209854P.
XX      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX      Meyer K, Van Dyk DE, Viitanen PV;
XX
XX      WPI; 2002-226795/28.
XX      P-PSDB; AAM50959.
XX
XX      Producing para-hydroxy benzoic acid in green plant, comprises expression
XX      of unique expression cassette containing gene encoding chorismate
XX      pyruvate lyase operably linked to specific chloroplast targeting
XX      sequence.
XX
XX      Claim 17; Page 57; 60pp; English.
XX
XX      The present nucleotide sequence is that of an open reading frame encoding
XX      chloroplast-targeted chorismate pyruvate lyase (CPL) fusion protein TP-
XX      CPL (see AAM50959). PCR (see ABA91841-42) was used to generate a DNA
XX      fragment corresponding to the transit peptide of tomato Rubisco small
XX      subunit and the first 4 amino acid residues of mature Rubisco. The DNA
XX      fragment was ligated into pET24a-CPL, which carries the Escherichia coli
XX      CPL open reading frame (see ABA91837). The TP-CPL construct is an example
XX      of expression cassettes of the invention that are designed for the high-
XX      level production of p-hydroxybenzoic acid (pHBA) in higher plants.
XX      Cleavage of TP-CPL in the chloroplast releases a novel polypeptide (see
XX      AAM50961) that has full enzyme activity, converting chorismate to pHBA. A
XX      claimed plant comprising a CPL expression cassette is selected from
XX      soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
XX      barley, oats, sorghum, rice, Arabidopsis, sugarcane, canola,
XX      millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
XX      component of liquid crystal polymers which have application in the
XX      automotive, electrical and other industries. (Updated on 29-AUG-2003 to
XX      standardise OS field)
XX
XX      SQ      Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2,04e-92      Length:      684
Score:           887.00      Matches:      170
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      100.0%      Indels:      0
DB:              6      Gaps:        0

US-10-718-311-16 (1-170) x ABA91838 (1-684)

Qy      1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db      172 ATGCAGGTGTGCGATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCGCTATTGT 231

Qy      21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db      232 AAAGAGATCCCTGCCCTGGATCCGCACTGCTCGACTGGCTGTGCTGGAGATTCCATG 291

Qy      41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db      292 ACMAAACGTTTGAACACAGCAGGAAACCGGTAAACGGTGACGATATCCGCGAGGGTTT 351

Qy      61 ValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db      352 GTCGACGAGATGAATCCCGCAAGAACTCCGCGCTGCTGCCGAAAGAGCTCGTTACTGG 411

Qy      81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100

```

```

Db      412 TTACGTGAATTTGTTATGTGCCGATGGTGAAACCGTGGCTTGCCTGTCGATCCGTCGTT 471
Qy      101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db      472 CCTGTGTCAACGTTAAGCGGCGCGAGCTGGCGTTACAAAAATTGGGTAAACGCGGTTA 531
Qy      121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db      532 GGACGCTATCTGTTACATCATCATGACATTAAACCGGGGACTTTATTGTAGATAGGCGGTGAT 591
Qy      141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db      592 GCCGGGCTGTGGGGCGACGTTCCCGCTCGCATTAACGGGTAAACCGCTGTTGCTAACA 651
Qy      161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db      652 GNACTGTTTTTACCAGCGGTACACCGTTGTAC 681

RESULT 2
ABK47712
ID      ABK47712 standard; DNA; 684 BP.
XX
XX      AC      ABK47712;
XX
XX      DT      18-JUN-2002 (first entry)
XX      DE      DNA sequence of ORF for TP-CPL fusion protein.
XX      KW      SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester; plant;
XX      KW      sinapoylglucose;malate sinapoyltransferase; glycosylated aromatic acid;
XX      KW      malate conjugated aromatic acid; polymer synthesis; ubiC; TP-CPL; gene;
XX      KW      carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;
XX      KW      tomato Rubisco small subunit precursor; rbcS2; mutant; ds.
XX      OS      Escherichia coli.
XX      OS      Lycopersicon esculentum.
XX      OS      Synthetic.
XX      OS      Chimeric.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      1..684
XX      FT      /*tag= a
XX      FT      /product= "TP-CPL fusion protein"
XX
XX      PN      WO200204653-A2.
XX
XX      PD      17-JAN-2002.
XX
XX      PF      05-JUL-2001; 2001WO-US021283.
XX
XX      PR      07-JUL-2000; 2000US-0216615P.
XX
XX      PA      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX      PI      Flint D, Meyer K, Viitanen PV;
XX
XX      WPI; 2002-303779/34.
XX      DR      P-PSDB; AAU77942.
XX
XX      PT      Producing aromatic acid conjugates, involves contacting glycosylated
XX      aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the
XX      presence of sinapoylglucose:malate sinapoyltransferase.
XX      Example; Page 70; 72pp; English.
XX
XX      The present invention relates to the isolation of Arabidopsis thaliana
XX      gene (SNG1, sinapoylglucose accumulator 1) encoding
XX      sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyzes the
XX      substitution of a glucose moiety on a glycosylated aromatic acid with a
XX      malate moiety to form a malate conjugated aromatic acid. The enzyme is
XX      useful for producing malate conjugated aromatic acids, carboxylic acid
XX      conjugated aromatic acids or aromatic esters. The malate conjugated

```

CC aromatic acids are useful in the synthesis of various polymers. The
CC present sequence representing the ORF (open reading frame) for TP-CPL
CC (tomato Rubisco small subunit precursor for rbcS2-Escherichia coli
CC chorismate pyruvate lyase (CPL)) fusion protein is described in the
CC examples of the present invention

XX Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,04e-92 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-10-718-311-16 (1-170) x ABK47712 (1-684)

QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGGCATATGTACACCCCGCGTTAAACGCAACTGCGTGGCTATTGT 231
QY 21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATG 291
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACANAAAGTTTTGAACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 351
QY 61 ValGluGlnAsnGluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTGAGCAGCAATGAATATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTG 411
QY 81 LeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTTGTATGTCCGATGGTGAACCGTGGCTGTGCCGCTGACCGTCGT 471
QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CTGTGTCAACGTTAAGCGGGCGGAGCTGGGTGTACAAAAATTTGGGTAAACCGCGTTA 531
QY 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAsp 140
Db 532 GCAGCGTATCTGTTCACATCATCATGACATTAAACCGGAGCTTTATTGAGATAGCGCTGAT 591
QY 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGGCTGTGGGGCGAGCTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTGTCTAACA 651
QY 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACGTGTTTTTACCGGCGTACCGCTGTGTATAC 681

RESULT 3

ADC25988

ID ADC25988 standard; DNA; 684 BP.

XX

AC ADC25988;

XX

DT 18-DEC-2003 (first entry)

DE Tomato/Escherichia coli chimeric TP-CPL DNA.

KW UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;
KW liquid crystal polymer; LCP; methylparaben; preservative; food;
KW cosmetic industry; ds; gene; tomato; TP-CPL; chimeric;
KW chorismate pyruvate lyase; plant.

XX Chimeric.

OS Lycopersicon esculentum.

OS Escherichia coli.

XX

FH Key Location/Qualifiers
CDS 1..684
FT /*tag= a
FT /product= "Tomato/Escherichia coli chimeric TP-CPL
FT protein"

XX WO2003066836-A2.

XX 14-AUG-2003.

XX 06-FEB-2003; 2003WO-US0005863.

XX 07-FEB-2002; 2002US-0355511P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Meyer K, Van Dyk DE, Viitanen PV;

XX WPI; 2003-767259/72.

XX P-PSDB; ADC25989.

PT New nucleic acid encoding UDP-glucosyltransferase, useful for preparing
PT cells that produce p-hydroxybenzoic acid glucose ester, also the new
PT enzymes.
XX Example 9; SEQ ID NO 41; 161pp; English.

XX The invention relates to a novel isolated nucleic acid that encodes a UDP
CC -glucosyltransferase. The method of the invention may be used to
CC transform microorganisms or green plant cells so that these produce a
CC higher level of high-volume chemicals or materials, such as PHBA (p-
CC hydroxybenzoic acid) ester glucoside which is an intermediate for PHBA, a
CC monomer for liquid crystal polymers (LCPs) and starting material for
CC methylparaben. Methylparaben is a preservative commonly used in the food
CC and cosmetic industries. The encoded enzymes may be used for in vitro
CC production of these compounds and for identifying similar enzymes by
CC sequence comparison. The current sequence is that of the
CC tomato/Escherichia coli chimeric TP-CPL (chorismate pyruvate lyase) DNA
CC of the invention.

XX Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,04e-92 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 10 Gaps: 0

US-10-718-311-16 (1-170) x ADC25988 (1-684)

QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGGCATATGTACACCCCGCGTTAAACGCAACTGCGTGGCTATTGT 231
QY 21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATG 291
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACANAAAGTTTTGAACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 351
QY 61 ValGluGlnAsnGluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTGAGCAGCAATGAATATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTG 411
QY 81 LeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTTGTATGTCCGATGGTGAACCGTGGCTGTGCCGCTGACCGTCGT 471
QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120

CC able to catalyze prodn. of an antiviral, bactericidal, fungicidal or
 CC insecticidal factor are new. A preferred gene is ubiC from E.coli
 CC (AAQ92409) which codes for chorismate-pyruvate lyase (AAR74742). The
 CC lyase catalyses conversion of chorismate to p-hydroxybenzoic acid and
 CC transgenic tobacco plants which express the active enzyme are resistant
 CC to tobacco mosaic virus
 XX
 SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.33e-89 Length: 495
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AAQ92409 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
 Db 1 ATGTCACACCCCGCGTTAAACGCACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 60
 Qy 26 LeuAspProGlnLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
 Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTCGGAGGATTCATGACAAAACGTTTGA 120
 Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
 Db 121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTTCGACGAGATGAA 180
 Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
 Db 181 ATCCCCGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGTTACGTGAATTTG 240
 Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
 Db 241 TTATGTCCGATGGTAACCGTGGTGTCCCGGTGTACCGTCTCTGTGTCAACGTTA 300
 Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
 Db 301 AGCGGCGGAGCTGGCGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTT 360
 Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 145
 Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCGCGTATGCCGGCTGTGGGG 420
 Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 Db 421 CGACGTTCCCGCTGCGATTAAAGCGGTAAACGCTGTGTTGCTAACGAACCTGTTTACCG 480
 Qy 166 AlaSerProLeuTyr 170
 Db 481 CGGTACACCGTTGTAC 495

RESULT 7

ABA91837

ID ABA91837 standard; DNA; 495 BP.

XX

AC ABA91837;

XX 15-MAY-2002 (first entry)

XX Escherichia coli chorismate pyruvate lyase gene coding region.

XX Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;

XX transgenic plant; ubiC; gene; ds.

XX Escherichia coli.

OS

XX Key Location/Qualifiers

XX 1..495

FT CDS /*tag= a

FT

FT /partial
 FT /product= "Chorismate pyruvate lyase"
 XX /note= "the CDS does not include a stop codon"

PN WO200194607-A2.

XX 13-DEC-2001.

XX 22-MAY-2001; 2001WO-US016661.

XX 02-JUN-2000; 2000US-0209854P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Meyer K, Van Dyk DE, Viitanen PV;

XX WPI; 2002-226795/28.

XX P-PSDB; AAM50958.

XX Producing para-hydroxy benzoic acid in green plant, comprises expression
 of unique expression cassette containing gene encoding chorismate
 pyruvate lyase operably linked to specific chloroplast targeting
 sequence.

XX Claim 4; Page 55; 60pp; English.

XX The present sequence is that of the coding region of the Escherichia coli
 strain W3110 ubiC gene in expression construct pET24a. The gene encodes
 chorismate pyruvate lyase (CPL, see AAM50968). It was obtained by PCR
 amplification of strain W3110 genomic DNA using primers (see ABA91839-40)
 based on the published E. coli ubiC gene. The CPL open reading frame is
 used in an expression cassette designed for the high-level production of
 p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette
 comprises the CPL coding sequence operably linked to a promoter capable
 of driving protein expression in higher plants. The cassette also has a
 sequence encoding a chloroplast transit peptide, its natural cleavage
 site, and a small portion of a transit peptide donor protein fused to the
 N-terminus of CPL. The chloroplast targeting sequence targets the foreign
 protein to the chloroplast and aids in its uptake into the organelle. The
 cleavage site is unique to the transit peptide, and cleavage of the
 chimeric protein encoded by the cassette at this site releases a novel
 polypeptide that has full enzyme activity, comprising the mature CPL
 enzyme and a small portion of transit peptide donor (see AAM50961). A
 plant comprising the CPL expression cassette is claimed, and may be
 soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
 barley, oats, sorghum, rice, Arabidopsis, sugarbeet, sugarcane, canola,
 millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
 component of liquid crystal polymers which have application in the
 automotive, electrical and other industries

XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

SQ

Alignment Scores:
 Pred. No.: 8.33e-89 Length: 495
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 6 Gaps: 0

US-10-718-311-16 (1-170) x ABA91837 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25

Db 1 ATGTCACACCCCGCGTTAAACGCAACTGCGTGGCTATTGTAAGAGATCCCTGCC 60

Qy 26 LeuAspProGlnLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45

Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTCGGAGGATTCATGACAAAACGTTTGA 120

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65

Db 121 CAGCAGGGAAGAAACGGTAAGCGGTACGATGATCCGGAAGGTTTGTTCGAGCAGATGAA 180

Qy	66	IleProGluGluLeuProLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAGAAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG	240
Qy	86	LeuCysAlaAaspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTACCGTCTCTGTCAAGTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	301	AGCGGCGGAGCTGGCGTTACAAAATGGGTAAAACGCCGTTAGACGCTATCTGTTTC	360
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCGCGTATGCCGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	421	CGAGCTTCCGCTCGGATTAAAGCGGTAACCGCTGTGTCTAACAGAACTGTTTTACCG	480
Qy	166	AlaSerProLeuTyr	170
Db	481	CGGTCACCGTTGTAC	495
RESULT 8			
ABK47709	ID	ABK47709 standard; DNA; 495 BP.	
XX	AC	ABK47709;	
XX	DT	18-JUN-2002 (first entry)	
XX	DE	DNA sequence of ORF for CPL for PET24a E. coli expression construct.	
XX	KW	SNGL; sinapoylglucose accumulator 1; SMT; aromatic ester;	
XX	KW	sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;	
XX	KW	malate conjugated aromatic acid; polymer synthesis; ubiC; CPL; gene;	
XX	KW	carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;	
XX	KW	PET24a; ds.	
XX	OS	Escherichia coli.	
XX	FH	Key	Location/Qualifiers
XX	FT	CDS	1..495
XX	FT		/*tag= a
XX	FT		/partial
XX	FT		/product= "CPL"
XX	FT		/note= "This sequence lacks a stop codon"
XX	PN	WO200204653-A2.	
XX	PN	17-JAN-2002.	
XX	XX		
XX	PF	05-JUL-2001; 2001WO-US021283.	
XX	XX	07-JUL-2000; 2000US-0216615P.	
XX	XX	(DUPO) DU PONT DE NEMOURS & CO E I.	
XX	PA	Flint D, Meyer K, Viitanen PV;	
XX	PI	WPI; 2002-303779/34.	
XX	DR	P-PSDB; AAU77941.	
XX	DR		
XX	XX		
XX	PT	Producing aromatic acid conjugates, involves contacting glycosylated	
XX	PT	aromatic acid with malate, malate-hydrocarboxylic acid or alcohol in the	
XX	PT	presence of sinapoylglucose:malate sinapoyltransferase.	
XX	XX		
XX	PS	Example; Page 68; 72pp; English.	
XX	XX		
CC	CC	The present invention relates to the isolation of Arabidopsis thaliana	
CC	CC	gene (SNGL, sinapoylglucose accumulator 1) encoding	

CC	sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the
CC	substitution of a glucose moiety on a glycosylated aromatic acid with a
CC	malate moiety to form a malate conjugated aromatic acid. The enzyme is
CC	useful for producing malate conjugated aromatic acids, carboxylic acid
CC	conjugated aromatic acids or aromatic esters. The malate conjugated
CC	aromatic acids are useful in the synthesis of various polymers. The
CC	present sequence representing the ORF (open reading frame) for chorismate
CC	pyruvate lyase (CPL) in pET24a Escherichia coli expression construct is
CC	described in the examples of the present invention
XX	
SQ	Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No.:	8.33e-89 Length: 495
Score:	854.00 Matches: 165
Percent Similarity:	100.0% Conservative: 0
Best Local Similarity:	100.0% Mismatches: 0
Query Match:	96.3% Indels: 0
DB:	6 Gaps: 0
US-10-718-311-16 (1-170) x ABK47709 (1-495)	
QY	6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysIysGluIleProAla 25
DB	1 ATGTCACACCCCGGCTTAACGCACTGCGTGCCTATTGTAAAGAGATCCCTGCC 60
QY	26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrIysArgPheGlu 45
DB	61 CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCCATGCACAAAACGTTTGAA 120
QY	46 GlnGlnGlyIysThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 65
DB	121 CAGCAGGGGAAAAACGGTAAGCGTGACGATGATCCCGAAGGGTTGTCTGAGCAGAATGAA 180
QY	66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
DB	181 ATCCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240
QY	86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
DB	241 TTATGTGCCGATGTGTGAACCGTGCGTTCGCCGTCGTACCGTCGTTCTGTGTCAAGTTA 300
QY	106 SerGlyProGluLeuAlaLeuGlnIysLeuGlyIysThrProLeuGlyArgTyrIlePhe 125
DB	301 AGCGGGCGGAGCTGGCGTTACAAAAATTGGGTAAACGCCGTTAGGACGCTATCTGTTC 360
QY	126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
DB	361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCCGGTATGCGGGGTGTGGGG 420
QY	146 ArgArgSerArgLeuArgLeuSerGlyIysProLeuLeuThrGluLeuPheLeuPro 165
DB	421 CGAGCTTCCCGCCTGCGGATTAAAGCGGTAAACCGCTGTGTGCTAACAGACTGTTTACCG 480
QY	166 AlaSerProLeuTyr 170
DB	481 GCGTCACCGTTGTAC 495
RESULT 9	
ADC25984	
ID	ADC25984 standard; DNA; 495 BP.
XX	
AC	ADC25984;
XX	
DT	18-DEC-2003 (first entry)
XX	
DE	Escherichia coli chorismate pyruvate lyase DNA.
XX	
KW	UDP-glucosyltransferase; pHBA; p-hydroxybenzoic acid ester glucoside;
KW	liquid crystal polymer; LCP; methylparaben; preservative; food;
KW	cosmetic industry; ds; gene; ubiC; CPL; chorismate pyruvate lyase.
XX	
OS	Escherichia coli.

XX FH Key Location/Qualifiers
 XX FT 1. .495
 XX FT /*tag= a
 XX FT /partial
 XX FT /product= "Escherichia coli chorismate pyruvate lyase
 XX FT protein"
 XX FT /note= "No stop codon"
 XX
 XX PN WO2003066836-A2.
 XX PD 14-AUG-2003.
 XX PF 06-FEB-2003; 2003WO-US005863.
 XX PR 07-FEB-2002; 2002US-0355511P.
 XX XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX PI Meyer K, Van Dyk DE, Viitanen PV;
 XX XX WPI; 2003-767259/72.
 XX DR P-PSDB; ABC25985.
 XX XX
 XX PT New nucleic acid encoding UDP-glucosyltransferase, useful for preparing
 XX FT cells that produce p-hydroxybenzoic acid glucose ester, also the new
 XX FT enzymes.
 XX XX
 XX PS Example 9; SEQ ID NO 37; 161pp; English.
 XX CC The invention relates to a novel isolated nucleic acid that encodes a UDP
 CC -glucosyltransferase. The method of the invention may be used to
 CC transform microorganisms or green plant cells so that these produce a
 CC higher level of high-volume chemicals or materials, such as pHBA (p-
 CC hydroxybenzoic acid) ester glucoside which is an intermediate for pHBA, a
 CC monomer for liquid crystal polymers (LCPs) and starting material for
 CC methylparaben. Methylparaben is a preservative commonly used in the food
 CC and cosmetic industries. The encoded enzymes may be used for in vitro
 CC production of these compounds and for identifying similar enzymes by
 CC sequence comparison. The current sequence is that of the Escherichia coli
 CC chorismate pyruvate lyase (CPL; UbiC) DNA of the invention.
 XX SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.33e-89 Length: 495
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 10 Gaps: 0

US-10-718-311-16 (1-170) x ABC25984 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
 Db 1 ATGTCACACCCGCGCTTAACGCACTGCGTGCCTGCTATTGTAAAGAGATCCCTGCC 60
 Qy 26 LeuAepProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
 Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAACGTTTGAA 120
 Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
 Db 121 CAGCAGGGGAAAACCGTGAAGCGTGACGATGATCCGGAAGGGTTTGTGAGCAGCAATGAA 180
 Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
 Db 181 ATCCCCGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGTTACGTGAATAATTTG 240
 Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
 Db 241 TTATGTCCGATGGTGAACCGTGGCTTGCCGGTGTGCTACCGTCTTCCTGTGTCAACGTTA 300

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuPhe 125
 Db 301 AGCGGGCGGAGCTGGCGTTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
 Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
 Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGCTGTGGGG 420
 Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 Db 421 CGAGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTGTCTAACAGAACTGTTTTACCG 480
 Qy 166 AlaSerProLeuTyr 170
 Db 481 CGGTACCGTTGTAC 495

RESULT 10
 ADV91618
 ID ADV91618 standard; DNA; 498 BP.
 XX AC
 XX AC ADV91618;
 XX DT 10-MAR-2005 (first entry)
 XX DE Escherichia coli chorismate pyruvate lyase (CPL) DNA.
 XX KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
 KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;
 KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
 KW neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid; gene;
 KW ds.
 XX OS Escherichia coli.
 FH Key Location/Qualifiers
 FT CDS 1. .498
 FT /*tag= a
 FT /product= "Escherichia coli CPL protein"
 XX US2004261147-A1.
 XX 23-DEC-2004.
 XX PF 16-JUN-2003; 2003US-00462162.
 XX PR 16-JUN-2003; 2003US-00462162.
 XX XX (MEYE/) MEYER K.
 XX PA (VIIT/) VIITANEN P V.
 XX PA (FLIN/) FLINT D.
 XX PI Meyer K, Viitanen PV, Flint D;
 XX WPI; 2005-057232/06.
 XX DR P-PSDB; ADV91619.
 XX PT Producing hydroquinone glucoside in a green plant comprises growing a
 PT green plant having nucleic acid fragments and chorismate pyruvate lyase
 PT expression cassette.
 XX Example 4; SEQ ID NO 29; 70pp; English.
 XX The present invention relates to methods and materials to produce
 CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
 CC in genetically modified green plants and microorganisms. The method
 CC relies upon transgenic plants or genetically modified microorganisms that
 CC produce increased levels of the initial substrate para-hydroxybenzoic
 CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
 CC arbutin. Plants and microbes can be genetically engineered to produce
 CC high levels of pHBA either by functional expression of the bacterial

CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
 CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
 CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
 CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
 CC present sequence is Escherichia coli chorismate pyruvate lyase (CPL) DNA.
 CC This sequence is found in the E. coli expression vector construct
 CC pET24a(+).
 XX

SQ Sequence 498 BP; 114 A; 122 C; 141 G; 121 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.4e-89 Length: 498
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 14 Gaps: 0

US-10-718-311-16 (1-170) x ADV91618 (1-498)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
 DB 1 ATGTACACCCCGCTTACGCAACTGCTGCGCTATGTAAAGAGATCCCTGCC 60
 QY 26 LeuAspProGlnLeuLeuAspTrrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
 DB 61 CTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120
 QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 65
 DB 121 CAGCAGGGAANAACGGTAAGCGTACGATCCGCAAGGGTGTTCGAGCAGAATGAA 180
 QY 66 IleProGluGluLeuProLeuProLysGluSerArgTyrTrrpLeuArgGluIleLeu 85
 DB 181 ATCCCGGAGAACTGCGCTGTCCGAAAGAGTCTGTTACTGTTACGTGAATTTG 240
 QY 86 LeuCysAlaAspGlyGluProTrrpLeuAlaGlyArgThrValProValSerThrLeu 105
 DB 241 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTCAACGTTA 300
 QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
 DB 301 AGCGGCGGAGCTGCGCTTACAAAATTTGGTAAACCGCTATAGGACGCTATCTGTT 360
 QY 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrrpGly 145
 DB 361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCCGCTGATGCCGGCTGTGGGG 420
 QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 DB 421 CGACGTTCCCGCTCGATTAAGCGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 480
 QY 166 AlaSerProLeuTyr 170
 DB 481 GCGTCACCGTTGAC 495

RESULT 11

ID AAD06956

XX AAD06956 standard; DNA; 854 BP.

XX AAD06956;

AC AAD06956;

XX 06-AUG-2001 (first entry)

DE ubiC gene encoding chorismate pyruvate lyase.

XX p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;
 KW chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;
 KW PEP synthase; tkkA; transketolase; aroC; chorismate synthase; aroA;
 KW 5-enolpyruvoylshikimate-3-phosphate synthase; BPS synthase; aroB;
 KW shikimate kinase II; aroB; 3-dehydroquinate synthase; BPS synthase; aroG;
 KW 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase; da.
 XX

OS Escherichia coli.

XX US6210937-B1.

XX 03-APR-2001.

XX 22-APR-1998; 98US-00064693.

XX 22-APR-1997; 97US-0044094P.

XX (BECH-) BECHTEL BWXT IDAHO LLC.

XX Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;

XX WPI; 2001-280857/29.

XX Producing aromatic compounds, especially commercially acceptable levels

XX of p-hydroxybenzoic acid, comprises developing genetically engineered

XX bacteria that carry selected genes of the common aromatic pathway.

XX Claim 5; Col 31-32; 25pp; English.

XX The present invention relates to a method of producing p-hydroxybenzoic

XX acid (pHB) which comprises transforming microorganisms with plasmids

XX carrying selected genes of the common aromatic pathway and by conversion

XX of carbon sources. The selected genes of the common aromatic pathway of

XX the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG

XX structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate

XX (DHAP) isoenzyme synthase (phe), tkkA gene encoding transketolase, ppsA

XX gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding

XX chorismate synthase, aroI gene encoding shikimate kinase II, aroA gene

XX encoding 5-enolpyruvoylshikimate-3-phosphate (BPS) synthase and aroB

XX gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for

XX converting chorismate to a selected aromatic compound. The method is

XX particularly used for the production of commercially acceptable levels of

XX aromatic compounds, particularly p-hydroxybenzoic acid. The present

XX sequence is ubiC gene which encodes chorismate pyruvate lyase that

XX catalyses the production of chorismate from pHB

XX

SQ Sequence 854 BP; 202 A; 200 C; 222 G; 230 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.78e-88 Length: 854

Score: 854.00 Matches: 165

Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0

Query Match: 96.3% Indels: 0

DB: 4 Gaps: 0

US-10-718-311-16 (1-170) x AAD06956 (1-854)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25

DB 321 ATGTACACCCCGCTTACGCAACTGCTGCGCTATGTAAAGAGATCCCTGCC 380

QY 26 LeuAspProGlnLeuLeuAspTrrpLeuLeuGluAspSerMetThrLysArgPheGlu 45

DB 381 CTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 440

QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 65

DB 441 CAGCAGGGAANAACGGTAAGCGTACGATGATCCGCAAGGGTGTTCGAGCAGAATGAA 500

QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrrpLeuArgGluIleLeu 85

DB 501 ATCCCGGAGAACTGCGCTGCTGCGAAAGAGTCTGTTACTGTTACGTAATTTG 560

QY 86 LeuCysAlaAspGlyGluProTrrpLeuAlaGlyArgThrValProValSerThrLeu 105

DB 561 TTATGTCCGATGGTGAACCGTGGCTTCCCGGTGCTGCTGTTCTGTGTCAACGTTA 620

QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125

Db 621 AGCGGGCGGAGTGGCGCTTACAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 680
 Qy 126 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
 Db 681 ACATCATCGACATTAAACCGGGACTTTATTAGATAGGCGCGTGTATGCCGGGCTGTGGGG 740
 Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 Db 741 CGACGTTCCCGCCTCGGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTACC 800
 Qy 166 AlaSerProLeuTyr 170
 Db 801 GCGTCACCGTTGTAC 815
 RESULT 12
 AAD06957
 ID AAD06957 standard; DNA; 6641 BP.
 XX
 AC AAD06957;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE pME2 comprising ubiC, aroG, tktA and ppsA genes of the aromatic pathway.
 XX
 KW p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;
 KW chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;
 KW PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;
 KW 5-enolpyruvoylshikimate-3-phosphate synthase; BPS synthase; aroL;
 KW shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;
 KW 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase;
 KW pME2; ds.
 XX
 OS Escherichia coli.
 XX
 FN US6210937-B1.
 XX
 PD 03-APR-2001.
 XX
 PF 22-APR-1998; 98US-00064693.
 XX
 PR 22-APR-1997; 97US-0044094P.
 XX
 PA (BECH-) BECHTEL BWXT IDAHO LLC.
 XX
 PI Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;
 XX
 DR WPI; 2001-280857/29.
 XX
 PT Producing aromatic compounds, especially commercially acceptable levels
 PT of p-hydroxybenzoic acid, comprises developing genetically engineered
 PT bacteria that carry selected genes of the common aromatic pathway.
 XX
 PS Claim 4; Col 31-40; 25pp; English.
 XX
 CC The present invention relates to a method of producing p-hydroxybenzoic
 CC acid (pHB) which comprises transforming microorganisms with plasmids
 CC carrying selected genes of the common aromatic pathway and by conversion
 CC of carbon sources. The selected genes of the common aromatic pathway of
 CC the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG
 CC structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate
 CC (DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppsA
 CC gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding
 CC chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene
 CC encoding 5-enolpyruvoylshikimate-3-phosphate (BPS) synthase and aroB
 CC gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for
 CC converting chorismate to a selected aromatic compound. The method is
 CC particularly used for the production of commercially acceptable levels of
 CC aromatic compounds, particularly p-hydroxybenzoic acid. The present
 CC sequence is pME2 plasmid which comprises ubiC, aroG, tktA and ppsA genes
 CC involved in the aromatic pathway
 XX
 SQ Sequence 6641 BP; 1567 A; 1698 C; 1853 G; 1523 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.08e-87 Length: 6641
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 4 Gaps: 0
 US-10-718-311-16 (1-170) x AAD06957 (1-6641)
 Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
 Db 321 ATGTCACACCCCGCGTTAAACGCACTGGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 380
 Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
 Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGCACAAAACGTTTTTGA 440
 Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
 Db 441 CAGCAGGGGAAAAACGGTAAGCGTGACGATGATCCGGAAGGGTTTGTGCGACGAATGAA 500
 Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
 Db 501 ATCCCCGAGACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 560
 Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
 Db 561 TTATGTCCGATGTTGAACCGTGGCTTGGCGGTGCTACCGCTGTTCTGTCTCAACGTTA 620
 Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
 Db 621 AGCGGGCGGAGTGGCGGTACAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTTTC 680
 Qy 126 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
 Db 681 ACATCATCGACATTAAACCGGGACTTTATTAGATAGGCGCGTGTATGCCGGGCTGTGGGG 740
 Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 Db 741 CGACGTTCCCGCCTCGGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTACC 800
 Qy 166 AlaSerProLeuTyr 170
 Db 801 GCGTCACCGTTGTAC 815
 RESULT 13
 ADV91643
 ID ADV91643 standard; DNA; 1971 BP.
 XX
 AC ADV91643;
 XX
 DT 10-MAR-2005 (first entry)
 XX
 DE C. parapsilosis pHB 1-H- E. coli CPL chimeric DNA.
 XX
 KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
 KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;
 KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
 KW neoplasm; pHB 1-hydroxylase; pHB; para-hydroxybenzoic acid;
 KW p-hydroxybenzoic acid; gene; ds.
 XX
 OS Candida parapsilosis.
 OS Escherichia coli.
 OS Chimeric.
 XX
 FH Location/Qualifiers
 CDS 4..501
 FT /*tag= a
 FT /product= "Escherichia coli CPL protein"
 FT /transl_except= (pos: 46..48, aa:Cys)

CC specifically the Rhodobacter capsulatus transformant MC9R/pRSFAC
XX Sequence 2000 BP; 444 A; 462 C; 538 G; 556 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 1,29e-87 Length: 2000
Score: 851.00 Matches: 164
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.4% Mismatches: 0
Query Match: 95.9% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AAT29820 (1-2000)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 380 ATGTCACACCCCGGCTTAACGCACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 439
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 440 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 499
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 500 CAGCAGGGAANAACGGTAAGCGTGACGATGATGCGCGAAGGGTTGTGACACAGATGA 559
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 560 ATCCCGGAAGAACTGCCGCTGTCGCAAGAGATCTCGTTACTGCTTACGTGAATTTTG 619
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 620 TTATGTCCGATGGTAACCGTGGCTTGCCGCTGCTACCGTCTGTCGTGTCACAGTTA 679
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 680 AGCGGGCCGAGCTGCGCTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTT 739
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 740 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGTGTCGGGG 799
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 800 CGAGTTCGCCCTCGGATTAAGCGGTAAACCGCTGTGCTTAACAGAACTGTTTTACCG 859
Qy 166 AlaSerProLeuTyr 170
Db 860 GCGTCACCGTGTAC 874
RESULT 15
AAS90335
ID AAS90335 standard; cDNA; 1207 BP.
XX AAS90335;
AC AAS90335;
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26139.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; ABG26148.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 26139; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1207 BP; 316 A; 265 C; 288 G; 338 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,62e-83 Length: 1207
Score: 811.00 Matches: 163
Percent Similarity: 98.2% Conservative: 1
Best Local Similarity: 97.6% Mismatches: 1
Query Match: 91.4% Indels: 2
DB: 5 Gaps: 0

US-10-718-311-16 (1-170) x AAS90335 (1-1207)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 652 ATGTCACACCCCGGCTTAACGCACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 711
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 712 CTGGATCCGCAACTGCTCGACTGGCTGTCTGCGAGGATTCATGACAAAACGTTTGA 771
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 772 CAGCAGGGAANAACGGTAAGCGTGACGATGATCGCGAAGGGTTGTGACAGAGATGA 831
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 832 ATCCCGAAGAACTGCGCTGCTGCCAAGAGTCTCGTACTGCTTACGTGCAATTTTG 891
Qy 86 -LeuCysAlaAspGlyGluProTrpLeuAla-GlyArgThrValProValSerThrL 105
Db 892 TGTATGTGCCGATGGCGAACCGCGGCTTGGCGGGTCTGACCGTCTCTGTGTCACAGT 951
Qy 105 euserGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuP 125
Db 952 TAAAGCGGGCCGAGCTGGCGTTACAAAATTTGGGTAAACGCCGCTTATCTGT 1011

```

QY 125 heThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 145
Db |||||
QY 1012 TCACATCATCGACATTAAACCGGACTTTATTGAGATAGGCCGTGATGCCGGCTGTGGG 1071
Db |||||
QY 145 lyArgArgSerArgLeuArgLeuSerClyIysProLeuLeuThrGluLeuPheLeuP 165
Db |||||
QY 1072 GCGGACGTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTAC 1131
Db |||||
QY 165 roAlaSerProLeuTyr 170
Db |||||
QY 1132 CGGCGTCACCGTTGTAC 1148
Db |||||

```

Search completed: May 30, 2006, 02:42:24
 Job time : 753.985 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2006, 20:59:50 ; Search time 725.015 Seconds
(without alignments)
2380.132 Million cell updates/sec

Title: US-10-718-311-4
Perfect score: 854
Sequence: 1 MSHPALQRLALRYCKEIPA.....RLSGKPLILITELFLPASPLY 165

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10718311/runat_26052006_164914_13428/app_query.fasta_1
-DB=N Geneseq -QFMT=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10718311 -CGEN_1_1_2295 -runat_26052006_164914_13428 -NCFU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_8.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*
15: Geneseqn2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

RESULT 1
AAQ92409
ID AAQ92409 standard; DNA; 495 BP.
XX
AC AAQ92409;
XX
DT 11-JAN-1996 (first entry)
XX
DE E.coli chorismate-pyruvate lyase coding sequence.
XX
KW Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;
KW tobacco mosaic virus resistance; antiviral activity;
KW p-hydroxybenzoic acid; ss.
OS Escherichia coli.
XX
FN DE4423022-C1.
XX
PD 24-MAY-1995.
XX
PF 30-JUN-1994; 94DE-04423022.
XX
PR 30-JUN-1994; 94DE-04423022.
XX
PA (HEID/) HEIDE L.
XX
PI Heide L, Siebert M, Severin K;

have this spans noted

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2006, 20:59:50 ; Search time 725.015 Seconds
(without alignments)
2380.132 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQTRALRYCKEIPA.....RLSGKPLLTLELFPASPLY 165

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/abs/ABSSWEB spool/US10718311/runat_26052006_164914_13428/app_query.fasta_1
-DB=N_Geneseq -QFM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p
-USER=US10718311 @CN 1 1 2295 @runat_26052006_164914_13428 -NCPUs=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.8.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*
- 15: Geneseqn2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	854	100.0	495	2	AaQ92409 E.coli ch
2	854	100.0	495	6	Aba91837 Escherich
3	854	100.0	495	6	Abk47709 DNA seque

4	854	100.0	495	10	ADC25984	Adc25984 Escherich
5	854	100.0	498	14	ADV91618	Adv91618 Escherich
6	854	100.0	684	6	ABA91838	Aba91838 Chloropla
7	854	100.0	684	6	ABK47712	Abk47712 DNA seque
8	854	100.0	684	10	ADC25988	Adc25988 Tomato/Es
9	854	100.0	684	14	ADV91622	Adv91622 Tomato ru
10	854	100.0	854	4	AAD06956	Aad06956 ubiC gene
11	854	100.0	3452	14	ADV91636	Adv91636 C. paraps
12	854	100.0	6641	4	AAD06957	Aad06957 pME2 comp
13	851	99.6	2000	2	AAT29820	Aat29820 E. coli u
14	843	98.7	1371	14	ADV91643	Adv91643 C. paraps
15	811	95.0	1207	5	AAS90335	Aas90335 DNA encod
16	790	92.5	1918	2	AAQ67671	Aaq67671 E.coli ub
17	673	78.8	645	11	ACH95439	Ach95439 Klebsiell
18	417	48.8	519	10	ACF70957	Acf70957 Photorhab
19	417	48.8	110000	10	ACF67367_40	Continuation (41 o
20	417	48.8	110000	10	ACF65388_07	Continuation (8 of
21	401	47.0	537	10	ADF02837	Adf02837 Bacterial
22	186	21.8	630	11	ABD08838	Abd08838 Pseudomon
23	174.5	20.4	519	12	ADL04203	Adl04203 DNA encod
24	174	20.4	100	8	ACD78861	Act78861 E. coli K
25	171.5	20.1	269223	4	AAF28554	Aaf28554 Genomic f
26	167	19.6	100	8	ACD78860	Act78860 E. coli K
27	165	19.3	100	8	ACD78859	Act78859 E. coli K
28	157.5	18.4	540	9	ADA29056	Ada29056 DNA encod
29	146.5	17.2	1248	11	ABD08757	Abd08757 Pseudomon
30	130.5	15.3	609	12	ADN36906	Adn36906 X. albili
31	130.5	15.3	55839	12	ADN36893	Adn36893 X. albili
32	121	14.2	195	10	ADF02801	Adf02801 Bacterial
33	97.5	11.4	475	6	AAS97252	Aas97252 Neisseria
34	90	10.5	2127	4	ABA95459	Aba95459 Thermus t
35	90	10.5	9133	3	AAA81730	Aaa81730 N. mening
36	90	10.5	110000	3	AAA81490_07	Continuation (8 of
37	90	10.5	34980	3	AAF21608	Aaf21608 Neisseria
38	88	10.3	42811	12	ADQ97948	Adq97948 Human can
39	87	10.2	59589	11	ACN44724	Acn44724 Mouse gen
40	86.5	10.1	110000	10	ACF67367_43	Continuation (44 o
41	86.5	10.1	110000	10	ACF65388_04	Continuation (5 of
42	86	10.1	2892	15	AEF63232	Aef63232 Human UNC
43	85.5	10.0	522	10	ABZ33133	Abz33133 N. gonorr
44	85.5	10.0	2157	1	AAH70438	Aan70438 Streptoco
45	85	10.0	25785	4	AAK78465	Aak78465 Human imm

ALIGNMENTS

RESULT 1
AAQ92409
ID AAQ92409 standard; DNA; 495 BP.

XX

AC AAQ92409;

XX

DT 11-JAN-1996 (first entry)

XX E.coli chorismate-pyruvate lyase coding sequence.

XX Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;
KW tobacco mosaic virus resistance; antiviral activity;
KW p-hydroxybenzoic acid; ss.

XX Escherichia coli.

XX DE4423022-C1.

XX 24-MAY-1995.

XX 30-JUN-1994; 94DE-04423022.

XX 30-JUN-1994; 94DE-04423022.

XX (HEID/) HEIDE L.

XX Heide L, Siebert M, Severin K;

have this
spms noted

XX WPI; 1995-186908/25.
 DR P-PSDB; AAR74742.
 XX Transgenic plants with increased content of resistance factor - contg.
 PT bacterial DNA coding for enzyme catalysing prodn. of resistance factor,
 PT e.g. antiviral p-hydroxy-benzoic acid.
 XX Claim 3; Page 5; 7pp; German.
 XX Transgenic plants containing a bacterial gene which encodes an enzyme
 CC able to catalyse prodn. of an antiviral, bactericidal, fungicidal or
 CC insecticidal factor are new. A preferred gene is ubiC from E.coli
 CC (AAQ92409) which codes for chorismate-pyruvate lyase (AAR74742). The
 CC lyase catalyses conversion of chorismate to p-hydroxybenzoic acid and
 CC transgenic tobacco plants which express the active enzyme are resistant
 CC to tobacco mosaic virus
 XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6.52e-89 Length: 495
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x AAQ92409 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
 Db 1 ATGTCACACCCCGCGCTTAACGCATCTGCGTGGCTGATTTAAAGAGATCCCTGCC 60
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
 Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGAA 120
 Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
 Db 121 CAGCAGGCGAAACCGTAAAGCGTGACGATGATCCGCAAGGGTTTGTTCAGCAGAAATGAA 180
 Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
 Db 181 ATCCCGGAAGAACTGCCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 240
 Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
 Db 241 TTATGTGCGGATGGTGAACCGTGGCTTSCCGTGGTACCGTCTCTGTGTCACCGTTA 300
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
 Db 301 AGCGGGCGGAGCTGGCGCTTACAAAATGGGTAAACGCCGTTAGGACGCTATCTGTTC 360
 Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
 Db 361 ACATCATCGACATTAAACCCGGAGCTTTATTGAGATAGCCGCTGTGCGGCTGTGGGG 420
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
 Db 421 CGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
 Db 481 CGGTACCGTGTATC 495

RESULT 2

ABA91837

ID ABA91837 standard; DNA; 495 BP.

XX ABA91837;

XX AC

DT 15-MAY-2002 (first entry)

XX Escherichia coli chorismate pyruvate lyase gene coding region.
 DE
 XX Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;
 KW transgenic plant; ubiC; gene; ds.
 XX Escherichia coli.
 OS
 XX Key Location/Qualifiers
 FH 1. .495
 CDS /*tag= a
 FT /partial
 FT /product= "Chorismate_pyruvate_lyase"
 FT /note= "the CDS does not include a stop codon"
 XX WO200194607-A2.
 XX 13-DEC-2001.
 XX 22-MAY-2001; 2001WO-US016661.
 XX 02-JUN-2000; 2000US-0209854P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Meyer K, Van Dyk DE, Viitanen PV;
 XX WPI; 2002-226795/28.
 XX P-PSDB; AAM50958.

Producing para-hydroxy benzoic acid in green plant, comprises expression

of unique expression cassette containing gene encoding chorismate
 pyruvate lyase operably linked to specific chloroplast targeting
 sequence.
 Claim 4; Page 55; 60pp; English.
 The present sequence is that of the coding region of the Escherichia coli
 strain W3110 ubiC gene in expression construct pET24a. The gene encodes
 chorismate pyruvate lyase (CPL, see AAM50968). It was obtained by PCR
 amplification of strain W3110 genomic DNA using primers (see ABA91839-40)
 based on the published E. coli ubiC gene. The CPL open reading frame is
 used in an expression cassette designed for the high-level production of
 p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette
 comprises the CPL coding sequence operably linked to a promoter capable
 of driving protein expression in higher plants. The cassette also has a
 sequence encoding a chloroplast transit peptide, its natural cleavage
 site, and a small portion of a transit peptide donor protein fused to the
 N-terminus of CPL. The chloroplast targeting sequence targets the foreign
 protein to the chloroplast and aids in its uptake into the organelle. The
 cleavage site is unique to the transit peptide, and cleavage of the
 chimeric protein encoded by the cassette at this site releases a novel
 polypeptide that has full enzyme activity, comprising the mature CPL
 enzyme and a small portion of transit peptide donor (see AAM50961). A
 plant comprising the CPL expression cassette is claimed, and may be
 soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
 barley, oats, sorghum, rice, Arabidopsis, sugarbeet, sugarcane, canola,
 millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
 component of liquid crystal polymers which have application in the
 automotive, electrical and other industries

Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6.52e-89 Length: 495
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 6 Gaps: 0

US-10-718-311-4 (1-165) x ABA91837 (1-495)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGCTTAACGCAACTCGCTGCGCTATTTGTAAGAGATCCCTGCC 60
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGCTGCTGCGGAGATTCATGACAAACGTTTGGAA 120
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGAACCGTAGCGTGACGATGATCCGCGAGGGTTTGTGCGACGAGATGAA 180
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCGCGAAAGAGTCTCGTTACTGTGTACGTGAAATTTG 240
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCCGATGGTAACCGTGCTGCGCGGTACCGTCTGCTGTCAACGTTA 300
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGCGAGTGGCTGTACAAAATTGGGTAAACCGCTTAGACGCTATCTGTTC 360
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGGTGATCGCGGCTGTGGGG 420
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTGCGATTAAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTACCG 480
QY 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495

RESULT 3

ID ABK47709 standard; DNA; 495 BP.
XX
AC ABK47709;
XX
DT 18-JUN-2002 (first entry)
XX
DE DNA sequence of ORF for CPL in pET24a E. coli expression construct.
XX
KW SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester;
KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;
KW malate conjugated aromatic acid; polymer synthesis; ubiC; CPL; gene;
KW carboxylic acid conjugated aromatic acid; chorisinate pyruvate lyase;
KW pET24a; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT 1. .495
FT CDS /tag= a
FT /partial
FT /product= "CPL"
FT /note= "This sequence lacks a stop codon"
XX
PN WO200204653-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021283.
XX
PR 07-JUL-2000; 2000US-0216615P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Flint D, Meyer K, Viitanen PV;
XX

DR WPI; 2002-303779/34.
DR P-PSDB; AAU77941.
XX
PT Producing aromatic acid conjugates, involves contacting glycosylated
PT aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the
PT presence of sinapoylglucose:malate sinapoyltransferase.
XX
PS Example; Page 68; 72pp; English.
XX
CC The present invention relates to the isolation of Arabidopsis thaliana
CC gene (SNG1, sinapoylglucose accumulator 1) encoding
CC sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the
CC substitution of a glucose moiety on a glycosylated aromatic acid with a
CC malate moiety to form a malate conjugated aromatic acid. The enzyme is
CC useful for producing malate conjugated aromatic acids, carboxylic acid
CC conjugated aromatic acids or aromatic esters. The malate conjugated
CC aromatic acids are useful in the synthesis of various polymers. The
CC present sequence representing the ORF (Open reading frame) for chorisinate
CC pyruvate lyase (CPL) in pET24a Escherichia coli expression construct is
CC described in the examples of the present invention
XX
SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.52e-89 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: Gaps: 0

US-10-718-311-4 (1-165) x ABK47709 (1-495)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGCTTAACGCAACTCGCTGCGCTATTTGTAAGAGATCCCTGCC 60
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGCTGCTGCGGAGATTCATGACAAACGTTTGGAA 120
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGAACCGTAGCGTGACGATGATCCGCGAGGGTTTGTGCGACGAGATGAA 180
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTGTAGTGAATTTG 240
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCCGATGGTAACCGTGCTGCGCGTGTACCGTCTCTGTGTCAACGTTA 300
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGCGAGTGGCTGTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 360
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGGTGATGCGCGGCTGTGGGG 420
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTGCGATTAAGCGGTAAACCGCTGTGCTTAACAGAACTGTTTTACCG 480
QY 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495
RESULT 4
ID ADC25984 standard; DNA; 495 BP.
XX

```

AC ADC25984;
XX
DT 18-DEC-2003 (first entry)
XX
DE Escherichia coli chorismate pyruvate lyase DNA.
XX
KW UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;
KW liquid crystal polymer; LCP; methylparaben; preservative; food;
KW cosmetic industry; ds; gene; ubiC; CPL; chorismate pyruvate lyase.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..495
FT /tag= a
FT /partial
FT /product= "Escherichia coli chorismate pyruvate lyase
FT protein"
FT /note= "No stop codon"
XX
XX WO2003066836-A2.
XX
XX 14-AUG-2003.
XX
XX 06-FEB-2003; 2003WO-US005863.
XX
XX 07-FEB-2002; 2002US-0355511P.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Meyer K, Van Dyk DE, Viitanen PV;
XX
XX WPI; 2003-767259/72.
XX
XX P-PSDB; ADC25985.
XX
XX New nucleic acid encoding UDP-glucosyltransferase, useful for preparing
XX cells that produce p-hydroxybenzoic acid glucose ester, also the new
XX enzymes.
XX
XX Example 9; SEQ ID NO 37; 161pp; English.
XX
XX The invention relates to a novel isolated nucleic acid that encodes a UDP
XX -glucosyltransferase. The method of the invention may be used to
XX transform microorganisms or green plant cells so that these produce a
XX higher level of high-volume chemicals or materials, such as PHBA (p-
XX hydroxybenzoic acid) ester glucoside which is an intermediate for PHBA, a
XX monomer for liquid crystal polymers (LCPs) and starting material for
XX methylparaben. Methylparaben is a preservative commonly used in the food
XX and cosmetic industries. The encoded enzymes may be used for in vitro
XX production of these compounds and for identifying similar enzymes by
XX sequence comparison. The current sequence is that of the Escherichia coli
XX chorismate pyruvate lyase (CPL; UbiC) DNA of the invention.
XX
XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.52e-89 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 10 Gaps: 0

US-10-718-311-4 (1-165) x ADC25984 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGCGTTAACGCACTGCTGGCTGCGCTATTGTTAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACACAAACGTTTGAA 120

```

Producing hydroquinone glucoside in a green plant comprises growing a green plant having nucleic acid fragments and chorismate pyruvate lyase expression cassette.

XX The present invention relates to methods and materials to produce
PS hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX in genetically modified green plants and microorganisms. The method
CC relies upon transgenic plants or genetically modified microorganisms that
CC produce increased levels of the initial substrate para-hydroxybenzoic
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
CC arbutin. Plants and microbes can be genetically engineered to produce
CC high levels of pHBA either by functional expression of the bacterial
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
CC present sequence is *Escherichia coli* chorismate pyruvate lyase (CPL) DNA.
CC This sequence is found in the *E. coli* expression vector construct
CC pET24a(+).

SQ Sequence 498 BP; 114 A; 122 C; 141 G; 121 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6,57e-89 Length: 498
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x ADV91618 (1-498)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
DB 1 ATGTACACCCCGGTAAACGACACTCGTGGCTATTTGAAGAGATCCCTGCC 60
QY 21 LeuAspProGlnLeuLeuAspTrrPLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
DB 61 CTGGATCCGCAACTGCTGACCTGGCTGTGTGGAGATTCCATGACAAACGTTTGA 120
QY 41 GlnGlnGlyValThrValSerValThrMetIleArgGluGlyPheValGlnHndGlu 60
DB 121 CAGCAGGGAAACCGTAAAGCGTGAAGCGATGATCCGCAAGGGGTTCGTCAGCAGATGAA 180
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
DB 181 ATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAATTTG 240
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
DB 241 TTATGTGCGCGATGTAACCGTGGCTGCGGTGCTGACCGTTCCTGTCTCAACGTTA 300
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
DB 301 AGCGGCGCGAGTGGCGGTACAAAATTTGGGTAAACCCGTTAGGACGCTATCTGTTTC 360
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
DB 361 ACATCATGCACATTAAACCGGACCTTATTGAGATAGCCGCGGTGTCGGGGG 420
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
DB 421 CGACGCTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480
QY 161 AlaSerProLeuTyr 165
DB 481 GCGTCACCGTTGAC 495

RESULT 6
ABA91838
ID ABA91838 standard; DNA; 684 BP.
XX
AC ABA91838;
XX

DT 29-AUG-2003 (revised)
XX 15-MAY-2002 (first entry)
DE Chloroplast transit peptide-chorismate pyruvate lyase fusion gene.
XX Chloroplast transit peptide; tomato; Rubisco; plant;
KW ribulose-1,5-bisphosphate carboxylase; chorismate pyruvate lyase; CPL;
KW enzyme; p-hydroxybenzoic acid; transgenic plant; gene; ds.
XX Lycopersicon esculentum.
OS *Escherichia coli*.
XX Chimeric.
FH Key Location/Qualifiers
CDS 1..495
FT /*tag= a
FT /product= "chloroplast-targeted CPL fusion"
FT transit_peptide 1..186
FT /*tag= b
FT /note= "tomato Rubisco chloroplast targeting sequence"
FT mat_peptide 187..492
FT /*tag= c
FT /note= "E. coli chorismate pyruvate lyase"
FT WO200194607-A2.
XX 13-DEC-2001.
PF 22-MAY-2001; 2001WO-US016661.
XX 02-JUN-2000; 2000US-0209854P.
PR (DUPO) DU PONT DE NEMOURS & CO E I.
XX Meyer K, Van Dyk DE, Viitanen PV;
PI WPI; 2002-226795/28.
DR P-PSDB; AAM50959.
XX Producing para-hydroxy benzoic acid in green plant, comprises expression
PT of unique expression cassette containing gene encoding chorismate
PT pyruvate lyase operably linked to specific chloroplast targeting
PT sequence.
XX Claim 17; Page 57; 60pp; English.
XX The present nucleotide sequence is that of an open reading frame encoding
CC chloroplast-targeted chorismate pyruvate lyase (CPL) fusion protein TP-
CC CPL (see AAM50959). PCR (see ABA91841-42) was used to generate a DNA
CC fragment corresponding to the transit peptide of tomato Rubisco small
CC subunit and the first 4 amino acid residues of mature Rubisco. The DNA
CC fragment was ligated into pET24a-CPL, which carries the *Escherichia coli*
CC CPL open reading frame (see ABA91837). The TP-CPL construct is an example
CC of expression cassettes of the invention that are designed for the high-
CC level production of p-hydroxybenzoic acid (pHBA) in higher plants.
CC Cleavage of TP-CPL in the chloroplast releases a novel polypeptide (see
CC AAM50961) that has full enzyme activity, converting chorismate to pHBA. A
CC claimed plant comprising a CPL expression cassette is selected from
CC soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
CC barley, oats, sorghum, rice, Arabidopsis, sugarcane, sugarcane, canola,
CC millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
CC component of liquid crystal polymers which have application in the
CC automotive, electrical and other industries. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,02e-88 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0

```

DB: 6 Gaps: 0
US-10-718-311-4 (1-165) x ABA91838 (1-684)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
DB 187 ATGTCACACCCCGCTTAACGCACTGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
DB 247 CTGGATCCGCAACTCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
DB 307 CAGCAGGGAAACCGTAAGCGTGACGATATCCGGAAGGTTTGTGAGAGAAATGAA 366
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
DB 367 ATCCCGGAAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
DB 427 TTATGTGCGCATGTTGAACCGTGGCTTCCGCTCGTACCGTCTGTTCTGTCAACGTTA 486
QY 101 SerGlyProGlnLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
DB 487 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
DB 547 ACATCATCGACATTAACCCCGGACTTTTATTGAGATAGCCCGTGTATCGTGTTC 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
DB 607 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
DB 667 GCGTCACCGTTGTAC 681

RESULT 7
ID ABK47712 standard; DNA; 684 BP.
XX AC ABK47712;
XX DT 18-JUN-2002 (first entry)
XX DE DNA sequence of ORF for TP-CPL fusion protein.
XX KW SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester; plant;
XX KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;
XX KW malate conjugated aromatic acid; polymer synthesis; ubiC; TP-CPL; gene;
XX KW carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;
XX KW tomato Rubisco small subunit precursor; rbcS2; mutant; ds.
XX OS Escherichia coli.
XX OS Lycopersicon esculentum.
XX OS Synthetic.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX CDS 1..684
XX FT /*tag= a
XX FT /product= "TP-CPL fusion protein"
XX PN WO200204653-A2.
XX PD 17-JAN-2002.
XX PF 05-JUL-2001; 2001WO-US021283.
XX PR 07-JUL-2000; 2000US-021661SP.

```

```

XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX Flint D, Meyer K, Viitanen PV;
XX WPI: 2002-303779/34.
XX P-PSDB; RAU77942.
XX Producing aromatic acid conjugates, involves contacting glycosylated
XX aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the
XX presence of sinapoylglucose:malate sinapoyltransferase.
XX Example; Page 70; 72pp; English.
XX The present invention relates to the isolation of Arabidopsis thaliana
XX gene (SNG1, sinapoylglucose accumulator 1) encoding
XX sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the
XX substitution of a glucose moiety on a glycosylated aromatic acid with a
XX malate moiety to form a malate conjugated aromatic acid. The enzyme is
XX useful for producing malate conjugated aromatic acids, carboxylic acid
XX conjugated aromatic acids or aromatic esters. The malate conjugated
XX aromatic acids are useful in the synthesis of various polymers. The
XX present sequence representing the ORF (open reading frame) for TP-CPL
XX (tomato Rubisco small subunit precursor for rbcS2-Escherichia coli
XX chorismate pyruvate lyase (CPL)) fusion protein is described in the
XX examples of the present invention
XX SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,02e-88 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0

US-10-718-311-4 (1-165) x ABK47712 (1-684)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
DB 187 ATGTCACACCCCGCTTAACGCAACTGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
DB 247 CTGGATCCGCAACTCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
DB 307 CAGCAGGGAAACCGTAAGCGTGACGATATCCGGAAGGTTTGTGAGAGAAATGAA 366
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
DB 367 ATCCCGGAAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
DB 427 TTATGTGCGCATGTTGAACCGTGGCTTCCGCTCGTACCGTCTGTTCTGTCAACGTTA 486
QY 101 SerGlyProGlnLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
DB 487 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
DB 547 ACATCATCGACATTAACCCCGGACTTTTATTGAGATAGCCCGTGTATCGTGTTC 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
DB 607 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
DB 667 GCGTCACCGTTGTAC 681

```

Db 667 GCGTCACCGTTGTAC 681

RESULT 8

ADC25988

ID ADC25988 standard; DNA; 684 BP.

XX

AC ADC25988;

XX

DT 18-DEC-2003 (first entry)

XX

DE Tomato/Escherichia coli chimeric TP-CPL DNA.

XX

KW UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;

KW liquid crystal polymer; LCP; methylparaben; preservative; food;

KW cosmetic industry; ds; gene; tomato; TP-CPL; chimeric;

KW chorismate pyruvate lyase; plant.

XX

OS Chimeric.

OS Lycopersicon esculentum.

OS Escherichia coli.

XX

Key Location/Qualifiers

CDS 1..684

FT /*tag= a

FT /product= "Tomato/Escherichia coli chimeric TP-CPL

FT protein"

XX

PN WO2003066836-A2.

XX

PD 14-AUG-2003.

XX

PF 06-FEB-2003; 2003WO-US005863.

XX

PR 07-FEB-2002; 2002US-0355511P.

XX

PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX

PI Meyer K, Van Dyk DE, Viitanen PV;

XX

DR WPI; 2003-767259/72.

XX

DR P-PSDB; ADC25989.

XX

PT New nucleic acid encoding UDP-glucosyltransferase, useful for preparing

PT cells that produce p-hydroxybenzoic acid glucose ester, also the new

PT enzymes.

XX

PS Example 9; SEQ ID NO 41; 161pp; English.

XX

CC The invention relates to a novel isolated nucleic acid that encodes a UDP

CC -glucosyltransferase. The method of the invention may be used to

CC transform microorganisms or green plant cells so that these produce a

CC higher level of high-volume chemicals or materials, such as PHBA (p-

CC hydroxybenzoic acid) ester glucoside which is an intermediate for PHBA, a

CC monomer for liquid crystal polymers (LCs) and starting material for

CC methylparaben. Methylparaben is a preservative commonly used in the food

CC and cosmetic industries. The encoded enzymes may be used for in vitro

CC production of these compounds and for identifying similar enzymes by

CC sequence comparison. The current sequence is that of the

CC tomato/Escherichia coli chimeric TP-CPL (chorismate pyruvate lyase) DNA

CC of the invention.

XX

SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

XX

Alignment Scores:

Pred. No.: 1.02e-88 Length: 684

Score: 854.00 Matches: 165

Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0

Query Match: 100.0% Indels: 0

DB: 10 Gaps: 0

US-10-718-311-4 (1-165) x ADC25988 (1-684)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20

DB 187 ATGTACACCCCGCGTTAAACGCAACTGCGTGGCTATTTGAAAGATCCCTGCC 246

QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40

DB 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATGACAAAACGTTTGA 306

QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60

DB 307 CAGCAGGGGAAACCGTAAGCGTAGCATCCGCGAAGGGTTTCTCGAGCAGAATGAA 366

QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80

DB 367 ATCCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 426

QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100

DB 427 TTATGTGCGGATGGTGAACCGTGGCTTGCCTGCTGCTACCGTCTCTGTCAACGTTA 486

QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyValArgThrPhe 120

DB 487 ACGGGCCGAGCTGCGCTTACAAAATGGGTAAACCCCGTTAGACGCTATCTGTTTC 546

QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140

DB 547 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCGCTGATCCCGGCTGTGGGG 606

QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160

DB 607 CGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 666

QY 161 AlaSerProLeuTyr 165

DB 667 GCGTCACCGTTGTAC 681

RESULT 9

ADV91622

ID ADV91622 standard; DNA; 684 BP.

XX

AC ADV91622;

XX

DT 10-MAR-2005 (first entry)

XX

DE Tomato rubisco small subunit precursor TP-E. coli CPL chimeric DNA.

XX

KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;

KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;

KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;

KW genetically engineered microorganism; antioxidant; antimicrobial agent;

KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;

KW neoplasm; PHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid;

KW rubisco small subunit precursor; transit peptide;

KW chloroplast transit peptide; TP; chloroplast; gene; ds.

OS Lycopersicon esculentum.

OS Escherichia coli.

OS Chimeric.

XX

Key Location/Qualifiers

CDS 1..684

FT /*tag= a

FT /product= "Tomato rubisco small subunit precursor TP-E.

FT coli CPL fusion protein"

XX

US2004261147-A1.

XX

PD 23-DEC-2004.

XX

PF 16-JUN-2003; 2003US-00462162.

XX

PR 16-JUN-2003; 2003US-00462162.

XX

```

PA (MEYE/) MEYER K.
PA (VIIT/) VIITANEN P V.
PA (FLIN/) FLINT D.
XX
XX Meyer K, Viitanen PV, Flint D;
XX
XX WPI; 2005-057232/06.
DR P-PSDB; ADV91623.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
PT green plant having nucleic acid fragments and chorismate pyruvate lyase
PT expression cassette.
XX
XX Example 4; SEQ ID NO 33; 70pp; English.
XX
XX The present invention relates to methods and materials to produce
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
CC in genetically modified green plants and microorganisms. The method
CC relies upon transgenic plants or genetically modified microorganisms that
CC produce increased levels of the initial substrate para-hydroxybenzoic
CC acid (p-hydroxybenzoic acid; PHBA) in a biosynthetic pathway to produce
CC arbutin. Plants and microbes can be genetically engineered to produce
CC high levels of PHBA either by functional expression of the bacterial
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
CC present sequence is tomato rubisco small subunit precursor (rbcS2).
CC transit peptide (chloroplast transit peptide; Tp)-Escherichia coli CPL
CC chimeric DNA. This sequence is present in the E. coli expression vector
CC construct pET24a-Tp-CPL.
XX
XX SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.02e-88 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x ADV91622 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 187 ATGTACACACCCGGCTTAACGCAACTGCGTGGCTATTGTAAGAGATCCCTGCC 246
Qy 21 LeuAspProGlnLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGTGAGAGGATTCATGACAAAACGTTTGAA 306
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGAANAACGGTAAGCGTGACGATATCCGCAAGGGTGTTCGACAGATGAA 366
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 80
Db 367 ATCCCGCAAGAACTGCCGCTGTCGCAAAAGAGTCTCGTTACTGCTTACGTGAATTTTG 426
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTCCGATGGTGAACCGTGGCTTCCCGGTGATCCGTCGTTCGTGTCAACGTTA 486
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 487 AGCGGGCGGAGCTGGCGTTACAAAATTTGGTAAACGCCGTTAGGACGCTATCTGTTTC 546
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCCGGGACTTTTATTGAGATAGCCGCTGATGCCGGGCTGTGGGG 606
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160

```

```

Db 607 CGAGGTTCCCGCCCTGCAGTAAACCGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 666
Qy 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 10
AAD06956
ID AAD06956 standard; DNA; 854 BP.
XX
XX AC AAD06956;
XX
XX DT 06-AUG-2001 (first entry)
XX
XX ubiC gene encoding chorismate pyruvate lyase.
XX
XX p-hydroxybenzoic acid; PHB; aromatic pathway; carbon source; ubiC;
XX chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;
XX PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;
XX 5-enolpyruvoylshikimate-3-phosphate synthase; EPSP synthase; aroL;
XX shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;
XX 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase; ds.
XX Escherichia coli.
XX
XX US6210937-B1.
XX
XX 03-APR-2001.
XX
XX 22-APR-1998; 98US-00064693.
XX
XX 22-APR-1997; 97US-0044094P.
XX
XX (BECH-) BECHTEL BWXT IDAHO LLC.
XX
XX Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;
XX WPI; 2001-280857/29.
XX
XX Producing aromatic compounds, especially commercially acceptable levels
XX of p-hydroxybenzoic acid, comprises developing genetically engineered
XX bacteria that carry selected genes of the common aromatic pathway.
XX
XX Claim 5; Col 31-32; 25pp; English.
XX
XX The present invention relates to a method of producing p-hydroxybenzoic
XX acid (pHB) which comprises transforming microorganisms with plasmids
XX carrying selected genes of the common aromatic pathway and by conversion
XX of carbon sources. The selected genes of the common aromatic pathway of
XX the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG
XX structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate
XX (DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppsA
XX gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding
XX chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene
XX encoding 5-enolpyruvoylshikimate-3-phosphate (EPSP) synthase and aroB
XX gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for
XX converting chorismate to a selected aromatic compound. The method is
XX particularly used for the production of commercially acceptable levels of
XX aromatic compounds, particularly p-hydroxybenzoic acid. The present
XX sequence is ubiC gene which encodes chorismate pyruvate lyase that
XX catalyses the production of chorismate from PHB
XX
XX SQ Sequence 854 BP; 202 A; 200 C; 222 G; 230 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.39e-88 Length: 854
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 4 Gaps: 0

```

```
US-10-718-311-4 (1-165) x AAD06956 (1-854)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 321 ATGTACACCCCGGTTAAGCAACTGCGTGGCTATTGTAAAGAGATCCCTGCC 380
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 381 CTGGATCCGCAACTGCTCGACTGGCTTGTCTGGAGGATTCATGACAAAACGTTTGA 440
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGlnGlnAsnGlu 60
Db 441 CAGCAGGGAACGATGACGCGTACGATATCCGGAAGGGTTTGTCCGAGCAGATGAA 500
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 501 ATCCCGAGAGAACTGCGGTGCTGCCGAAAGAGTCTCGTTACTGGTTAGTGAATTTTG 560
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 561 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTGTGTCAACGTTA 620
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 621 AGCGGGCCGAGCTGCGGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 680
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 681 ACATCATCAGACATTACCCGGGACTTTATTGAGATAGGCGGTGATCCGGGCTGTGGGG 740
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 741 CGACGTCTCCGCGCTCGGATTAAGCGGTAAACCGCTGTGTGTAACAGAACTGTTTTTACC 800
QY 161 AlaSerProLeuTyr 165
Db 801 CGGTACCGTGTGAC 815

RESULT 11
ADV91636
ID ADV91636 standard; DNA; 3452 BP.
XX
AC ADV91636;
XX
DT 10-MAR-2005 (first entry)
XX
DE C. parapsilosis pHB 1-H- E. coli CPL- A. thaliana UGT72B1 chimeric DNA.
XX
KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
KW genetically engineered microorganism; antioxidant; antimicrobial agent;
KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
KW neoplasm; pHBA 1-hydroxylase; pHBA 1-H; pHBA; para-hydroxybenzoic acid;
KW p-hydroxybenzoic acid; UDP-glucosyltransferase; UGT72B1; gene; ds.
XX
OS Arabidopsis thaliana.
OS Candida parapsilosis.
OS Escherichia coli.
OS Chimeric.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 4..501
FT /tag= a
FT /product= "Escherichia coli CPL protein"
FT 541..1980
FT /tag= b
FT /product=
FT /transl_except= (pos: 559..561, aa:Trp)
FT 2004..3446
FT /tag= c
FT /product= "Arabidopsis thaliana UGT72B1 protein"
FT
XX
```

```
PN US2004261147-A1.
XX
PD 23-DEC-2004.
XX
PF 16-JUN-2003; 2003US-00462162.
XX
PR 16-JUN-2003; 2003US-00462162.
XX
PA (MEYE/) MEYER K.
XX
PA (VIIT/) VIITANEN P V.
PA (FLIN/) FLINT D.
XX
PI Meyer K, Viitanen PV, Flint D;
XX
WPI: 2005-057232/06.
DR P-PSDB; ADV91605, ADV91619, ADV91631.
DR
XX
PT Producing hydroquinone glucoside in a green plant comprises growing a
PT green plant having nucleic acid fragments and chorismate pyruvate lyase
PT expression cassette.
XX
PS Example 7; SEQ ID NO 47; 70pp; English.
XX
CC The present invention relates to methods and materials to produce
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
CC in genetically modified green plants and microorganisms. The method
CC relies upon transgenic plants or genetically modified microorganisms that
CC produce increased levels of the initial substrate para-hydroxybenzoic
CC acid (p-hydroxybenzoic acid; PHBA) in a biosynthetic pathway to produce
CC arbutin. Plants and microbes can be genetically engineered to produce
CC high levels of PHBA either by functional expression of the bacterial
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The
CC present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) -
CC Escherichia coli chorismate pyruvate lyase (CPL) and Arabidopsis thaliana
CC UDP-glucosyltransferase (UGT72B1) chimeric DNA.
XX
SQ Sequence 3452 BP; 932 A; 730 C; 856 G; 934 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.61e-88 Length: 3452
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x ADV91636 (1-3452)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 4 ATGTACACCCCGGTTAAGCAACTGCGTGGCTATTGTAAAGAGATCCCTGCC 63
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 64 CTGGATCCGCAACTGCTCGACTGGCTTGTCTGGAGGATTCATGACAAAACGTTTGA 123
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGlnGlnAsnGlu 60
Db 124 CAGCAGGGAACGATGACGCGTACGATATCCGGAAGGGTTTGTCCGAGCAGATGAA 183
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 184 ATCCCGAGAGAACTGCGGTGCTGCCGAAAGAGTCTCGTTACTGGTTAGTGAATTTTG 243
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 244 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTGTGTCAACGTTA 303
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
```

Db 304 AGCGGGCGGAGCTGGCGTTACAAAAATTTGGTAAACCGCGTTAGGACGCTATCTGTTC 363
 Qy 121 ThrSerSerThrLeuThrArgAspPheleGluileGlyArgAspAlaGlyLeuTrpGly 140
 Db 364 ACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGGTGTCGGCGCTGTGGGG 423
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
 Db 424 CGAGCTTCCCGCTCGGATTAAAGCGGTAAACCGCTTTTGTCTAACAGAACTGTTTTACCG 483
 Qy 161 AlaSerProLeuTyr 165
 Db 484 CGGTACCGTTGTAC 498
 RESULT 12
 AAD06957
 ID AAD06957 standard; DNA; 6641 BP.
 XX
 AC AAD06957;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE pME2 comprising ubiC, aroG, tktA and ppsA genes of the aromatic pathway.
 XX
 KW p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;
 KW chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;
 KW PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;
 KW 5-enolpyruvylshikimate-3-phosphate synthase; PEP synthase; aroL;
 KW shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;
 KW 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase;
 KW pME2; ds.
 XX
 OS Escherichia coli.
 XX
 PN US6210937-B1.
 XX
 PD 03-APR-2001.
 XX
 PF 22-APR-1998; 98US-00064693.
 XX
 PR 22-APR-1997; 97US-0044094P.
 XX
 PA (BECH-) BECHTEL BWXT IDAHO LLC.
 XX
 PI Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;
 XX
 DR WPI; 2001-280857/29.
 XX
 PT Producing aromatic compounds, especially commercially acceptable levels
 PT of p-hydroxybenzoic acid, comprises developing genetically engineered
 PT bacteria that carry selected genes of the common aromatic pathway.
 XX
 PS Claim 4; Col 31-40; 25pp; English.
 XX
 CC The present invention relates to a method of producing p-hydroxybenzoic
 CC acid (pHB) which comprises transforming microorganisms with plasmids
 CC carrying selected genes of the common aromatic pathway and by conversion
 CC of carbon sources. The selected genes of the common aromatic pathway of
 CC the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG
 CC structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate
 CC (DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppsA
 CC gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding
 CC chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene
 CC encoding 5-enolpyruvylshikimate-3-phosphate (PEP) synthase and aroB
 CC gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for
 CC converting chorismate to a selected aromatic compound. The method is
 CC particularly used for the production of commercially acceptable levels of
 CC aromatic compounds, particularly p-hydroxybenzoic acid. The present
 CC sequence is pME2 plasmid which comprises ubiC, aroG, tktA and ppsA genes
 CC involved in the aromatic pathway
 XX
 SQ Sequence 6641 BP; 1567 A; 1698 C; 1853 G; 1523 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2.38e-87 Length: 6641
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 4 Gaps: 0
 US-10-718-311-4 (1-165) x AAD06957 (1-6641)
 Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
 Db 321 ATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 380
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
 Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTCTGGAGGATTCATGACAAAACGTTTTGAA 440
 Qy 41 GlnGlnGlyLysThrValSerValThrMetileargGluGlyPheValGlnHnGlu 60
 Db 441 CAGCAGGGAAACCGTAAGCGTCACGATGATCCGGAAGGGTTTGTCCGACAGAAATGAA 500
 Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80
 Db 501 ATCCCCGAGAACTGCGCGCTGCTCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 560
 Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
 Db 561 TTATGTGCGCATGTTGAACCGTGGCTTCCCGCTGCTACCGTCTTCTGTGTCAACGTTA 620
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
 Db 621 AGCGGGCGGAGCTGGCGCTTACAAAATTTGGTAAACCGCGCTTAGGACGCTATCTGTTC 680
 Qy 121 ThrSerSerThrLeuThrArgAspPheleGluileGlyArgAspAlaGlyLeuTrpGly 140
 Db 681 ACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 740
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
 Db 741 CGAGCTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800
 Qy 161 AlaSerProLeuTyr 165
 Db 801 CGGTACCGTTGTAC 815
 RESULT 13
 AAT29820
 ID AAT29820 standard; DNA; 2000 BP.
 XX
 AC AAT29820;
 XX
 DT 17-JAN-1997 (first entry)
 XX
 DE E. coli ubiquinone biosynthetic enzyme genes ubiC and ubiA.
 KW Ubiquinone-10; biosynthetic enzyme; ubiC; ubiA; gene;
 KW photosynthetic bacteria; recombinant production; plasmid pRSFAC;
 KW glutamate synthase; gene promoter; Rhodobacter capsulatus; transformant;
 KW MCS9/pRSFAC; ss.
 XX
 OS Escherichia coli.
 XX
 Key Location/Qualifiers
 CDS 380..877
 FT /*tag= a
 FT /label= ubiC
 CDS 890..1762
 FT /*tag= b
 FT /label= ubiA
 FT 1868..1887
 FT repeat_unit
 FT /*tag= c
 FT /rpt_type= inverted

```

FT repeat_unit 1892..1911
FT /*tag= d
FT /*rpt_type= inverted
PN JP08107789-A.
XX
XX
XX 30-APR-1996.
XX
XX 13-OCT-1994; 94JP-00273071.
XX
XX 13-OCT-1994; 94JP-00273071.
XX
XX (ALPH-) ALPHA SHOKUHIN KK.
XX
XX WPI; 1996-262599/27.
XX P-PSDB; AAR97746, AAR97747.
XX
XX Genes coding for ubiquinone biosynthetic enzymes - useful for prodn. of
XX ubiquinone-10 by transformed photosynthetic bacteria.
XX
XX Claim 1; Fig 2; l1pp; Japanese.
XX
XX The present sequence comprises the ubiquinone biosynthetic enzyme genes
XX ubiC and ubiA, which were cloned from the chromosomal DNA of E. coli
XX (Kohara map phage DNA bank IF8 (634); Cell 50, 495-508 (1987)). A large
XX amt. of ubiquinone-10 can be produced by culturing photosynthetic
XX bacteria transformed with the novel plasmid pRSPAC, which contains the
XX ubiC and ubiA genes and the glutamate synthase gene promoter.
XX Specifically the Rhodobacter capsulatus transformant MC9R/pRSPAC
XX
XX Sequence 2000 BP; 444 A; 462 C; 538 G; 556 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1e-87 Length: 2000
Score: 851.00 Matches: 164
Percent Similarity: 100.0% Conservatives: 1
Best Local Similarity: 99.4% Mismatches: 0
Query Match: 99.6% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x AAT29820 (1-2000)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db ATGTCACACCGCGCTTACGCAACTGCGCTGCGCTATTGTAAGAGATCCCTGCC 439
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGAA 499
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db CAGCAGGGGAAACCGGTAAAGCTGACGATGATGCGGGAAGGGTTGTGCGAGCAGATGAA 559
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db ATCCCGGAAGAACTGCGCGTGTGCGGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 619
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db TTATGTGCGGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 679
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db ACCGGCGCGAGCTGGCTTACAAAATTGGGTAAACCGCTTGGACGCTATCTGTTTC 739
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db ACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 799
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 859

```

```

QY 161 AlaSerProLeuTyr 165
Db 860 GCGTCACCGTGTGTAC 874

RESULT 14
ADV91643
ID ADV91643 standard; DNA; 1971 BP.
XX
XX AC ADV91643;
XX
XX 10-MAR-2005 (first entry)
XX
XX C. parapsilosis pHB 1-H- E. coli CPL chimeric DNA.
XX
XX Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
XX chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
XX HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
XX genetically engineered microorganism; antioxidant; antimicrobial agent;
XX anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
XX neoplasm; pHBA 1-hydroxylase; pHBA; para-hydroxybenzoic acid;
XX p-hydroxybenzoic acid; gene; ds.
XX
XX Candida parapsilosis.
XX OS Escherichia coli.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX CDS 4..501
XX /*tag= a
XX /product= "Escherichia coli CPL protein"
XX /transl_except= (pos: 46..48, aa:Cys)
XX 522..1561
XX /*tag= b
XX /product= "Candida parapsilosis pHBA 1-H protein"
XX /transl_except= (pos: 540..542, aa:Trp)
XX
XX US2004261147-A1.
XX
XX 23-DEC-2004.
XX
XX 16-JUN-2003; 2003US-00462162.
XX
XX 16-JUN-2003; 2003US-00462162.
XX
XX (MEYE/) MEYER K.
XX (VIIT/) VIITANEN P V.
XX (FLIN/) FLINT D.
XX
XX Meyer K, Viitanen PV, Flint D;
XX
XX WPI; 2005-057232/06.
XX
XX P-PSDB; ADV91605, ADV91619.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
XX green plant having nucleic acid fragments and chorismate pyruvate lyase
XX expression cassette.
XX
XX Example 7; SEQ ID NO 54; 70pp; English.
XX
XX The present invention relates to methods and materials to produce
XX hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX in genetically modified green plants and microorganisms. The method
XX relies upon transgenic plants or genetically modified microorganisms that
XX produce increased levels of the initial substrate para-hydroxybenzoic
XX acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
XX arbutin. Plants and microbes can be genetically engineered to produce
XX high levels of pHBA either by functional expression of the bacterial
XX protein chorismate pyruvate lyase (CPL) or by the expression of 4-
XX hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
XX Arbutin is useful as an antioxidant, antimicrobial agent, anti-
XX inflammatory agent and inhibitor of carcinogenesis (melanoma). The
XX present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) and

```

```
CC Escherichia coli chorismate pyruvate lyase (CPL) chimeric DNA.
XX
SQ Sequence 1971 BP; 550 A; 383 C; 484 G; 554 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8, 23e-87 Length: 1971
Score: 843.00 Matches: 164
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 98.7% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x ADV91643 (1-1971)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 4 ATGTCACACCCCGCGTTAAACGCACTGCGTGGCTGCTATTTTAAAGAGATCCCTGCC 63

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 64 CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAACGTTTGA 123

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 124 CAGCAGGGAACCGTTAAGCGTGACGATGATCCGCAAGGTTTGTCTGACGACAGATGA 183

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 184 ATCCCGCAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 243

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 244 TTATGTGCCGATGGGAACCGTGGCTGCCGCTGATCCGCTGTTCCGTGTCAACGTTA 303

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 304 AGCGGCGCGAGCTGGCGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTT 363

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 364 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGCGTGATGCCGGCTGTGGGG 423

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 424 CGACGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACC 483

Qy 161 AlaSerProLeuTyr 165
Db 484 GCGTCACCGTTGTAC 498

RESULT 15
AAS90335
ID AAS90335 standard; cDNA; 1207 BP.
XX
AC AAS90335;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26139.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX P-PSDB; ABG26148.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 26139; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS84197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1207 BP; 316 A; 265 C; 288 G; 338 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2, 06e-83 Length: 1207
Score: 811.00 Matches: 163
Percent Similarity: 98.2% Conservative: 1
Best Local Similarity: 97.8% Mismatches: 1
Query Match: 95.0% Indels: 2
DB: 5 Gaps: 0

US-10-718-311-4 (1-165) x AAS90335 (1-1207)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 652 ATGTCACACCCCGCGTTAAACGCACTGCGTGGCTATTTGTAAGAGATCCCTGCC 711

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 712 CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAACGTTTGA 771

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 772 CAGCAGGGAACCGTTAAGCGTGACGATGATCCGCAAGGTTTGTCTGACGACAGATGA 831

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 832 ATCCCGCAAGAACTGCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 891

Qy 81 -LeuCysAlaAspGlyGluProTrpLeuAla-GlyArgThrValProValSerThrL 100
Db 892 TGTATGTCCGATGGCGAACC CGCGCTTGC CGGCTCGTACCGTCTGTTCTGTGTCAACGT 951

Qy 100 euSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuP 120
Db 952 TAACGGCGCGGAGCTGCGGTTACAAAATTGGGTAACACCGCGTATAGGACGCTATCTGT 1011
```

QY 120 heThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 140
Db |||||
1012 TCACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCCGTGATGCCGGCTGTGGG 1071
QY 140 lyArgArgSerArgLeuArgLeuSerGlyIysProLeuLeuLeuThrGluLeuPheLeup 160
Db |||||
1072 GCGGACGTTCCGCGCTGCGATTAAAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTAC 1131
QY 160 roAlaSerProLeuTyr 165
Db |||||
1132 CGCGGTCACCGTTGTAC 1148

Search completed: May 30, 2006, 02:42:29
Job time : 730.015 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2006, 01:30:02 ; Search time 5000.73 Seconds
(without alignments)

3164.934 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQALRYCKEIPA.....RLSGKPLLTLELFPASPLY 165

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh
-Q=abs/ABSWEB spoof/US10718311/runat_26052006_164916_13439/app_query.fasta_1
-DB=GenEmbl -OFT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes03h
-USER=US10718311 @CGN_1_1_5767 @runat_26052006_164916_13439 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_env.*
- 2: gb_pat.*
- 3: gb_ph.*
- 4: gb_pl.*
- 5: gb_pr.*
- 6: gb_ro.*
- 7: gb_sts.*
- 8: gb_sy.*
- 9: gb_un.*
- 10: gb_vi.*
- 11: gb_ov.*
- 12: gb_htg.*
- 13: gb_in.*
- 14: gb_om.*
- 15: gb_ba.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	854	100.0	495	2	A47932 Sequence 1
2	854	100.0	495	2	AR428733 Sequence
3	854	100.0	495	2	AR455409 Sequence

4	854	100.0	495	2	AX329364	AX329364 Sequence
5	854	100.0	495	2	AX382254	AX382254 Sequence
6	854	100.0	684	2	AR428736	AR428736 Sequence
7	854	100.0	684	2	AR455412	AR455412 Sequence
8	854	100.0	684	2	AX329368	AX329368 Sequence
9	854	100.0	684	2	AX382258	AX382258 Sequence
10	854	100.0	854	2	AR144788	AR144788 Sequence
11	854	100.0	1498	15	ECUBIA	E11274 ubiC and ub
12	854	100.0	2000	2	E11274	M91136 Escherichia
13	854	100.0	2000	15	ECUBICA	M94113 Escherichia
14	854	100.0	2000	15	ECUBIPLS	X57434 E.coli ubiC
15	854	100.0	2034	15	ECUBIAC	X66619 E.coli gene
16	854	100.0	2348	15	ECUBI	AR144789 Sequence
17	854	100.0	6641	2	AR144789	Continuation (43 o
18	854	100.0	110000	15	U00096_42	Continuation (43 o
19	854	100.0	110000	15	AP009048_42	U00006 E. coli chr
20	854	100.0	176195	15	ECUW89_42	Continuation (52 o
21	850	99.5	110000	15	AE005174_51	Continuation (51 o
22	846	99.1	110000	15	BA000007_50	Continuation (52 o
23	846	99.1	110000	15	BA000007_51	DO087228 Escherich
24	843	98.7	1396	15	DO087228_	Continuation (44 o
25	843	98.7	110000	15	AE005674_43	Continuation (35 o
26	843	98.7	110000	15	AE014073_34	Continuation (42 o
27	843	98.7	110000	15	CP000036_41	Continuation (43 o
28	839	98.2	110000	15	CP000034_42	Continuation (45 o
29	835	97.8	110000	15	CP000038_44	Continuation (48 o
30	833	97.5	110000	15	AE014075_47	E07849 gDNA encodi
31	798	93.4	1918	2	E07849	AE008898 Salmonell
32	725	84.9	23880	15	AE008898	Continuation (42 o
33	725	84.9	110000	15	CP000026_41	Continuation (43 o
34	725	84.9	110000	15	CP000026_42	Continuation (43 o
35	721	84.4	110000	15	AE014613_42	Continuation (44 o
36	721	84.4	110000	15	AE017220_43	AL627282 Salmonell
37	721	84.4	234050	15	AL627282	AR384505 Sequence
38	673	78.8	645	2	AR384505	AE013658 Yersinia
39	437.5	51.2	10668	15	AE013658	Continuation (5 of
40	437.5	51.2	110000	15	BX936398_04	AJ414142 Yersinia
41	437.5	51.2	214050	15	AJ414142	AE017128 Yersinia
42	437.5	51.2	290803	15	AE017128	Continuation (7 of
43	424.5	49.7	110000	15	EX950851_06	Continuation (13 o
44	417	48.8	349107	15	EX951873	Continuation (13 o
45	408.5	47.8	110000	15	EX897699_12	

ALIGNMENTS

RESULT 1	A47932	Sequence 1 from Patent WO9600788.	495 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A47932	Sequence 1 from Patent WO9600788.				
DEFINITION	A47932	Sequence 1 from Patent WO9600788.				
ACCESSION	A47932	Sequence 1 from Patent WO9600788.				
VERSION	A47932.1	GI:2301795				
KEYWORDS		Escherichia coli				
SOURCE		Escherichia coli				
ORGANISM		Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.				
REFERENCE		1 (bases 1 to 495)				
AUTHORS		Heide, L., Severin, K. and Siebert, M.				
TITLE		TRANSGENIC PLANTS WITH AN INCREASED SECONDARY SUBSTANCE CONTENT				
JOURNAL		Patent: WO 9600788-A 1 11-JAN-1996;				
COMMENT		HEIDE LUTZ (DE)				
FEATURES		Other publication AU 2879295 960125				
source		Location/Qualifiers				
		1. .495				
		/organism="Escherichia coli"				
		/mol_type="unassigned DNA"				
		/strain="MC4100"				
		/db_xref="taxon:562"				
ORIGIN						
Alignment Scores:						

```
Pred. No.: 1.54e-81 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x A47932 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGCGTTAACGCAACTGGCTGCGCTGCTATTGTAAAGAGATCCCTGCC 60

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGAAAAACGGTAAGCGTACCATGATCGCGAAGGGTTTGTGACAGAAATTTG 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAAACGGTAAGCGTACCATGATCGCGAAGGGTTTGTGACAGAAATTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCCGTACCGTCTGCTGTTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGGAGCTGGCGTTACAAAAATTGGTAAACGCCGTTAGGACGCTAICTGTT 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCGCGTGTGGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGTTTCCCGCTCGGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495

RESULT 3
AR455409
LOCUS AR455409 495 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 3 from patent US 6683231.
ACCESSION AR455409
VERSION AR455409.1 GI:42689957
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Meyer,K., Viitanen,P.V. and Van Dyk,D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 3 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES
source
1..495
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores: 1.54e-81 Length: 495
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2

US-10-718-311-4 (1-165) x AR455409 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGCGTTAACGCAACTGGCTGCGCTGCTATTGTAAAGAGATCCCTGCC 60

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGAAAAACGGTAAGCGTACCATGATCGCGAAGGGTTTGTGACAGAAATTTG 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCCGCTGCTCGCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCCGTACCGTCTGCTGTTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGGAGCTGGCGTTACAAAAATTGGTAAACGCCGTTAGGACGCTAICTGTT 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCGCGTGTGGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGTTTCCCGCTCGGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495

RESULT 2
AR428733
LOCUS AR428733 495 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 12 from patent US 6642036.
ACCESSION AR428733
VERSION AR428733.1 GI:40188463
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Flint,D., Meyer,K. and Viitanen,P.
TITLE Sinapoylglucose:malate sinapoyltransferase form malate conjugates
JOURNAL Patent: US 6642036-A 12 04-NOV-2003;
E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES
source
1..495
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores: 1.54e-81 Length: 495
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Gaps: 0
DB: 2

US-10-718-311-4 (1-165) x AR455409 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
```

	Db	61	CTGGATCCGC	AAC	TGCTCG	ACTGG	CTGTT	GTCTGG	AGG	ATTCC	ATG	CAAAA	CGTTT	GGAA	120								
Qy		41	GlnGlnGly	Ly	Thr	Val	Ser	Val	Thr	Met	Ile	Arg	Glu	Gly	Phe	Val	Glu	Gln	Asn	Glu	60		
Db		121	CAGCAGG	AA	AA	AAC	CGT	AAG	CGT	GAC	GAT	GAT	CCG	CA	AGG	GGT	TTG	TG	CAG	CAG	180		
Qy		61	Ile	Pro	Glu	Glu	Leu	Pro	Leu	Pro	Lys	Glu	Ser	Arg	Tyr	Trp	Leu	Arg	Glu	Ile	Leu	80	
Db		181	ATCCCCG	AA	GA	AACTG	CCG	CTG	TCG	CCG	AA	GAG	TCT	CGT	TACT	CGG	TAC	TG	TAC	GTG	AA	240	
Qy		81	Leu	Cys	Ala	Asp	Gly	Glu	Pro	Trp	Leu	Ala	Gly	Arg	Thr	Val	Val	Pro	Val	Ser	Thr	Leu	100
Db		241	TTATGTG	CCG	AT	GTTG	TGA	ACCCG	TGG	CGT	CGT	AC	CGT	CGT	CGT	CGT	CGT	CGT	CGT	CGT	CGT	300	
Qy		101	Ser	Gly	Pro	Glu	Leu	Ala	Leu	Gln	Lys	Leu	Gly	Lys	Thr	Pro	Leu	Gly	Arg	Tyr	Leu	Phe	120
Db		301	AGCGGG	CG	GAG	CTGG	CGT	TAC	AA	AA	AA	TTGG	GTA	AA	CG	CCG	TG	TAG	G	ACG	CTAT	CTG	360
Qy		121	Thr	Ser	Ser	Thr	Leu	Thr	Arg	Asp	Phe	Ile	Glu	Ile	Gly	Arg	Asp	Ala	Gly	Leu	Trp	Gly	140
Db		361	ACATCAT	CG	AT	TAA	CCG	GG	ACT	TTT	ATT	TG	AG	TAG	CG	CGT	GAT	CG	CGG	CTG	TG	GGG	420
Qy		141	Arg	Arg	Ser	Arg	Leu	Arg	Leu	Ser	Cly	Lys	Pro	Leu	Leu	Thr	Glu	Leu	Phe	Leu	Pro	160	
Db		421	CG	AGT	TCC	CCG	CTCG	GAT	TAA	CGG	GTA	AA	CGG	TG	TG	CTT	AG	CAG	AA	CTG	T	T	480
Qy		161	Ala	Ser	Pro	Leu	Tyr	165															
Db		481	CGGTC	AC	CGT	TGT	GAT	495															

RESULT 5

AX382254

LOCUS

495 bp

DNA

linear

PAT 18-MAR-200

```

DEFINITION Sequence 12 from Patent WO0204653.
ACCESSION AX382254
VERSION AX382254.1 GI:19577034
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
1
AUTHORS Flint,D., Meyer,K. and Viitanen,P.V.
TITLE Sinapolyglucose:malate sinapyltransferase form malate conjugates
from benzoic acid glucosides
JOURNAL Patent: WO 0204653-A 12 17-JAN-2002;
E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES
    source Location/Qualifiers
    1..495 /organism="Escherichia coli"
    /mol_type="unassigned DNA"
    /db_xref="taxon:562"

ORIGIN

Alignment Scores:
Pred. No.: 1.54e-81 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x AX382254 (1-495)

```

Accession	Protein	Length
1	MerSerHisproA1aLeuThrGluLeuAargAlaLeuAargTyrCysLysGluIleProAla	20
Qy		
1	ATGTCAACCCCGCGTTAAACCAACTGCTGCGCTATTGTAAAGAGATCCCTGCC	60
Db		
21	LeuAspProGlnLeuLeuAspTrrLeuLeuLeuGluAspSerMetThrLysAargPheGlu	40
Qy		
61	CTGGATCCGCAACTGCTCGACTGCTGTTTCTGGAGGATTCATGACAAAACGTTTTGAA	120
Db		

```
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTTCGACAGAAATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTACTGTTACGTGAATTTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 241 TTATGTGCGGATGGTGAACCGTGGCTTCCCGGTACCGTGGTTCCTGTGTCAACGTTA 300
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgThrLeuPhe 120
Db 301 AGCGGCGCGAGCTGGCGCTTACAAAATTGGGTAAACCGCGGTAGGACGCTATCTGTTC 360
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCGGACTTATTAGATAGGCGCGTGTGCGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGACGTTCCCGCTCGCATTAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495
RESULT 6
LOCUS AR428736 684 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 16 from patent US 6642036.
ACCESSION AR428736
VERSION AR428736.1 GI:40188466
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 684)
AUTHORS Flint,D., Meyer,K. and Viitanen,P.
TITLE Sinapoylgucose:malate sinapoyltransferase form malate conjugates
JOURNAL from benzoic acid glucosides
Patent: US 6642036-A 16 04-NOV-2003;
E. I. Du Pont de Nemours and Company; Wilmington;
WOX;
FEATURES
source Location/Qualifiers
Pred. No.: 1. .684
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 2.19e-81 Length: 684
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2
US-10-718-311-4 (1-165) x AR428736 (1-684)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTCACACCCCGCGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGAA 306
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTTCGACAGAAATGAA 366
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTTCGACAGAAATTTG 426
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
```

```
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAAGAAACGGTGCCTGCCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 426
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTGCGGATGGTGAACCGTGGCTTCCCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 486
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgThrLeuPhe 120
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCGTGTAGGACGCTATCTGTTC 546
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTATTAGATAGGCGCGTGTGCGGCTGTGGGG 606
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGACGTTCCCGCTCGCATTAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTTACCG 666
Qy 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681
RESULT 7
LOCUS AR455412 684 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 7 from patent US 6683231.
ACCESSION AR455412
VERSION AR455412.1 GI:42689960
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 684)
AUTHORS Meyer,K., Viitanen,P.V. and Van Dyk,D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 7 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES
source Location/Qualifiers
Pred. No.: 1. .684
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 2.19e-81 Length: 684
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2
US-10-718-311-4 (1-165) x AR455412 (1-684)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTCACACCCCGCGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGAA 306
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTTCGACAGAAATGAA 366
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAAGAAACGGTGCCTGCCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 426
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
```

```

Db      427 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCGTCTCTGTGTCAACGTTA 486
Qy      101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db      487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTATCTGTTC 546
Qy      121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db      547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGCGTGATGCCGGGCTGTGGGG 606
Qy      141 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db      547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGCGTGATGCCGGGCTGTGGGG 606
Qy      141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuLeuLeuLeuPheLeuPro 160
Db      607 CGACGTTCCCGCTGGCATTAAGCGGTAAACCGCTGTGTCTAACAGAACTGTTTTACCG 666
Qy      161 AlaSerProLeuTyr 165
Db      667 GCGTCACCGTTGTAC 681

RESULT 8
AX329368
LOCUS      AX329368      684 bp      DNA      linear      PAT 08-JAN-2002
DEFINITION Sequence 7 from Patent WO0194607.
ACCESSION AX329368
VERSION    AX329368.1 GI:18102367
KEYWORDS   .
SOURCE     synthetic construct
ORGANISM   other sequences; artificial sequences.
REFERENCE  1
AUTHORS    Meyer, K., van Dyk, D.B. and Viitanen, P.V.
TITLE      High level production of p-hydroxybenzoic acid in green plants
JOURNAL    Patent: WO 0194607-A 7 13-DEC-2001;
           E.I. Dupont De Nemours (US)
FEATURES   Location/Qualifiers
            source
              1..684
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="synthetic CPL"
ORIGIN
Alignment Scores:
Pred. No.:      2.19e-81      Length:      684
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:    0
DB:            2      Gaps:      0

US-10-718-311-4 (1-165) x AX329368 (1-684)
Qy      1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db      187 ATGTACACACCCCGCTTAACGCACTGCGTGGCGTATTGTAAAGAGATCCCTGCC 246
Qy      21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db      247 CTGGATCCGCAACTGCTCGACTGGCTGTCTGGAGGATTCATGACAAAAGCTTTTGA 306
Qy      41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60
Db      307 CAGCAGGGAACAAACGCTAAGCGTCGATGATCCGGAAGGGTTTGTCTGAGCAGAAATGA 366
Qy      61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db      367 ATCCCCGGAAGAACTGCCCGTGTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426
Qy      81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db      427 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTCTGCTTACTGTTACGTGAAATTTTG 486
Qy      101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120

```

```

Db      487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTC 546
Qy      121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db      547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGCGTGATGCCGGGCTGTGGGG 606
Qy      141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuLeuLeuPheLeuPro 160
Db      607 CGACGTTCCCGCTGGCATTAAGCGGTAAACCGCTGTGTCTAACAGAACTGTTTTACCG 666
Qy      161 AlaSerProLeuTyr 165
Db      667 GCGTCACCGTTGTAC 681

RESULT 9
AX382258
LOCUS      AX382258      684 bp      DNA      linear      PAT 18-MAR-2002
DEFINITION Sequence 16 from Patent WO0204653.
ACCESSION AX382258
VERSION    AX382258.1 GI:19577037
KEYWORDS   .
SOURCE     synthetic construct
ORGANISM   other sequences; artificial sequences.
REFERENCE  1
AUTHORS    Flint, D., Meyer, K. and Viitanen, P.V.
TITLE      Sinapolyglucose:malate sinapolytransferase form malate conjugates
           from benzoic acid glucosides
JOURNAL    Patent: WO 0204653-A 16 17-JAN-2002;
           E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES   Location/Qualifiers
            source
              1..684
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="open reading frame of the chloroplast-targeted CPL
                fusion protein"
ORIGIN
Alignment Scores:
Pred. No.:      2.19e-81      Length:      684
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:    0
DB:            2      Gaps:      0

US-10-718-311-4 (1-165) x AX382258 (1-684)
Qy      1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db      187 ATGTACACACCCCGCTTAACGCACTGCGTGGCGTATTGTAAAGAGATCCCTGCC 246
Qy      21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db      247 CTGGATCCGCAACTGCTCGACTGGCTGTCTGGAGGATTCATGACAAAAGCTTTTGA 306
Qy      41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60
Db      307 CAGCAGGGAACAAACGCTAAGCGTCGATGATCCGGAAGGGTTTGTCTGAGCAGAAATGA 366
Qy      61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db      367 ATCCCCGGAAGAACTGCCCGTGTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426
Qy      81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db      427 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTCTGCTTACTGTTACGTGAAATTTTG 486
Qy      101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db      487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTC 546

```



```
/protein_id="AA17028.1"  
/db_xref="GI:347889"  
/translation="NEWSITONKLLAFHRLMRDTPKIGALLLLWPTLWALWVATPGVPLQWIIAVFVAGVWLMRAGCVNDYADRFKFDGHVKRT"
```

ORIGIN

Alignment Scores:	
Pred. No.:	5,13e-81
Score:	854.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	15
	Length: 1498
	Matches: 165
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0
US-10-718-311-4 (1-165) x ECU081A (1-1498)	

```

VERSION M93136.1 GI:148099
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB
          gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
          gene.
SOURCE   Escherichia coli W3110
ORGANISM Escherichia coli W3110
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Lightner, V.A., Bell, R.M. and Modrich, P.
TITLE The DNA sequences encoding plbB and dgk loci of Escherichia coli
JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)
PUBMED 6309817
REFERENCE 2 (bases 1 to 2000)
AUTHORS Nichols, B.P. and Green, J.M.
TITLE Cloning and Sequencing of Escherichia coli ubiC and purification of
          chorismate lyase
JOURNAL Unpublished (1992)
COMMENT Original source text: Escherichia coli (sub_strain W3110, strain
          K-12) (library: Kohare lambda-1F8) DNA.
          Location/Qualifiers
            1..2000
              /organism="Escherichia coli W3110"
              /mol_type="genomic DNA"
              /strain="K-12"
              /sub_strain="W3110"
              /db_xref="taxon:316407"
              /map="91.8 min"
              /tissue_lib="Kohare lambda-1F8"
            380..877
              /gene="ubiC"
              /gene="ubiC"
              /gene="ubiC"
              /function="enzymatic chorismate --> p-hydroxybenzoate +
pyruvate"
          /experiment="experimental evidence, no additional details
          recorded"
          /citation=[2]
          /codon_start=1
          /transl_table=11
          /product="chorismate lyase"
          /protein_id="AAA24711.1"
          /db_xref="GI:148100"
          /translation="MSHPALTQLRALRYCKEIPALDQLLDWLLLESMTKRPQOQK
TVSVTMREGVFEQNEIPEDPLPKESRYWLREILLCADGEPWLAKRTVPVSTLSG
PELALQKLKGTPLGRYUFTSLTRDFEIGRDAGLWRRSRLSLGKPLLLTFLFLP
ASPLY"
            890..1762
              /gene="ubiA"
              /gene="ubiA"
              /gene="ubiA"
              /function="enzymatic - 3-octaprenyl-4-hydroxybenzoate
synthesis"
          /experiment="experimental evidence, no additional details
          recorded"
          /citation=[2]
          /codon_start=1
          /transl_table=11
          /product="4-hydroxybenzoate-octaprenyl transferase"
          /protein_id="AAA24712.1"
          /db_xref="GI:148101"
          /translation="MWSLTONKLLAFHRLMRTDKPTGALLLLWPTLWLVATPGVP
QLWILAVFVAGWLMRAAGCVNDYADKRFQGHVKRTANRPLPSGAVTEREARALFVV
LVLISFLVLTNTMTLLSIAALAWVPFMKRYTHLPQVILGAAPFGWSIPWAFAA
VSEVPKSLMFLNAILWAVDYQVAMVDRDDDKIGIKSTAILFLGQYDLIIGIL
QIGVLAUMAIIGELNGLGWYWSILVAGALFVYQOKLIANREREACFKAPMNNYVG
LVFLGLAWSVWFH"
            complement(1917..2000)
              /gene="plbB"
              /gene="plbB"
              /gene="plbB"
              /EC_number="2.3.1.15"
              /function="lysophosphatidic acid synthesis"

/experiment="experimental evidence, no additional details
recorded"
/citation=[1]
/codon_start=1
/transl_table=11
/product="sn-glycerol-3-phosphate acyltransferase"
/protein_id="AAA24713.1"
/db_xref="GI:148102"
/translation="MKVIQLLAEILITSDVRLTIESATQGE"

ORIGIN
Alignment Scores:
Pred. No.: 7,02e-81 Length: 2000
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 15 Gaps: 0

US-10-718-311-4 (1-165) x ECOUBICA (1-2000)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgfyrCysLysGluileProAla 20
Db 380 ATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 439
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 440 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGAGGATTCATGACAAAAACGTTTGAA 499
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 500 CAGCAGGGAACCGTTAAGCGTGACGTGACGTGATGTCGCAAGGGTTTGTGAGCAGAAATGAA 559
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgfyrTrpLeuArgGluileLeu 80
Db 560 ATCCCGCAGAACTGCGCGTGTGCGCAAGAGTCTGTTACTGCTTACGTGAAATTTTG 619
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 620 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTACCGTCGTTCCTGTGTCAACGTTA 679
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgfyrLeuPhe 120
Db 680 AGCGGGCGGAGCTGGGGTTACAAAATTTGGTAAAACCGCGTTAGGACGCTATCTGTTT 739
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAspAlaGlyLeuTrpGly 140
Db 740 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGCGTGATGCCGGCTGTGGGG 799
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 800 CGAGTTTCCCGCTGGCATTAAAGCGGTAAACCGCTGTGCTTACAGAACTGTTTTTACCG 859
Qy 161 AlaSerProLeuTyr 165
Db 860 GCGTCACCGTTGTAC 874

RESULT 14
ECOUBIPLS 2000 bp DNA linear BCT 26-APR-1993
LOCUS Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA)
DEFINITION gene complete cds, chorismate lyase (ubiC) gene complete cds,
          sn-glycerol-3-phosphate acyltransferase (plbB) gene, 3' end.
          M93413
ACCESSION M93413.1 GI:148106
VERSION 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB
KEYWORDS gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
          gene.
SOURCE Escherichia coli W3110
ORGANISM Escherichia coli W3110
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2000)

```

AUTHORS Lightner,V.A., Bell,R.M. and Modrich,P.
 TITLE The DNA sequences encoding plsB and dgk loci of Escherichia coli
 JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)
 PUBMED 6309817
 REFERENCE 2 (bases 1 to 2000)
 AUTHORS Nichols,B.P. and Green,J.M.
 TITLE Cloning and sequencing of Escherichia coli ubiC and purification of chorismate lyase
 JOURNAL J. Bacteriol. 174 (16), 5309-5316 (1992)
 PUBMED 1644758
 COMMENT Original source text: Escherichia coli (sub_strain W3110, strain K-12) (library: Kohara lambda) DNA.
 FEATURES
 source
 1..2000
 /organism="Escherichia coli W3110"
 /mol_type="genomic DNA"
 /strain="K-12"
 /sub_strain="W3110"
 /db_xref="taxon:316407"
 /map="91.8 min"
 /tissue_lib="Kohara lambda"
 380..877
 /gene="ubiC"
 380..877
 /gene="ubiC"
 /function="enzymatic chorismate --> p-hydroxybenzoate + pyruvate"
 /experiment="experimental evidence, no additional details recorded"
 /note="N-terminal amino acid sequence of chorismate lyase :NH2-SHPALTLQRLALRYCEIPAL-"
 /citation=[2]
 /codon_start=1
 /transl_table=11
 /product="chorismate lyase"
 /protein_id="AAA24716.1"
 /db_xref="GI:148107"
 /translation="MSHPALTLQRLALRYCKEIPALDPQLDMLLEDSTMTRKEFOQGGK TVSVTWIRGFEQNEIPELPLPKESRYWLEIILCADGEPWLAGRTVPVSTLSG PELALQKLGKTPLGRLFTSSLTTRDFEIGRDAGLWGRSRRLSLGKELLTLTELPF ASPLY"
 890..1762
 /gene="ubiA"
 890..1762
 /gene="ubiA"
 /function="enzymatic - 3-octaprenyl-4-hydroxybenzoate synthesis"
 /experiment="experimental evidence, no additional details recorded"
 /citation=[2]
 /number=2
 /codon_start=1
 /transl_table=11
 /product="4-hydroxybenzoate-octaprenyl transferase"
 /protein_id="AAA24717.1"
 /db_xref="GI:148108"
 /translation="MEWSLTQNKLLAFHRLMRTDKPIGALLLLWPTLWALWATPGVP QLWILAVFAGVWLMRAGCVNDYADKFDGKVRXTANRPLPSGAVTEKEARALFV LVLSFLLVLTNTWLTILLSIALALAWYPMKRYTHLPQVVLGAAGFWSIPMAFAA VSESFPLSCWLMFLANILWAVDYOYAMVDKDDVKIGIKSTAILFGYDKLIIGIL QTVLALMAIIGELNGLWGYYSILVAGALFVYQOKLIANRERERACPKAFMNNYVG LVFLGLAMSYWHF"
 complement(1917..2000)
 /gene="plsB"
 complement(1917..2000)
 /gene="plsB"
 /EC_number="2.3.1.15"
 /function="lysophosphatidic acid synthesis"
 /experiment="experimental evidence, no additional details recorded"
 /citation=[1]
 /codon_start=1
 /transl_table=11

/product="sn-glycerol-3-phosphate acyltransferase"
 /protein_id="AAA24718.1"
 /db_xref="GI:148109"
 /translation="MKVYQLLAELITSDVRLTIESATQSEG"
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,02e-81 Length: 2000
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservatives: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 15 Gaps: 0
 US-10-718-311-4 (1-165) x ECOUBIPLS (1-2000)
 QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysIysGluIleProAla 20
 |||||
 Db 380 ATGTCAACCCCGGTTAAACGCAACTGCGTGCCTATTGTAAAGAGATCCCTGCC 439
 QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrIysArgPheGlu 40
 |||||
 Db 440 CTGGATCCGCAACTGCTCGACTGGCTGTGGAGAGATTCATGACAAACGTTTTGAA 499
 QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
 |||||
 Db 500 CAGCAGGGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGGTTTGTCCGAGCAGATGAA 559
 QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
 |||||
 Db 560 ATCCCCGAAAGACTGCGCTGCTGCGCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 619
 QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
 |||||
 Db 620 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679
 QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
 |||||
 Db 680 AGCGGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACCCCGTTAGGACGCTATCTGTTTC 739
 QY 121 ThrSerSerThrIleuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
 |||||
 Db 740 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGCGTGATCGCGGCTGTGGGG 799
 QY 141 ArgAspSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
 |||||
 Db 800 CCACGTTCCCGCTCGGATTAGCGGTAACCGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 859
 QY 161 AlaSerProLeuTyr 165
 |||||
 Db 860 GCGTCACCGTGTGAC 874
 RESULT 15
 ECUBIAC
 LOCUS
 DEFINITION E.coli ubiC and ubiA genes for chorismate lyase and 2034 bp DNA linear BCT 05-OCT-1992
 X57434
 ACCESSION X57434
 VERSION X57434.1 GI:43233
 KEYWORDS 4-hydroxybenzoate octaprenyl transferase; 4-hydroxybenzoate synthetase; chorismate lyase.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCE 1
 AUTHORS Nishimura,K., Nakahigashi,K. and Inokuchi,H.
 TITLE Location of the ubiA gene on the physical map of Escherichia coli
 JOURNAL J. Bacteriol. 174 (17), 5762 (1992)
 PUBMED 1512213
 REFERENCE 2 (bases 1 to 2034)
 AUTHORS Nishimura,K.
 TITLE Direct Submission

JOURNAL Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto University, Sakyo-ku, Kyoto 606, Japan

FEATURES Location/Qualifiers

source 1..2034

/organism="Escherichia coli"

/mol_type="genomic DNA"

/strain="K12 W3110"

/db_xref="taxon:562"

/map="91.7 min"

/clone="634"

/clone_lib="Kohara library"

-35_signal 137..142

-10_signal 159..164

gene 291..899

/gene="ubiC"

CDS 291..899

/gene="ubiC"

/codon_start=1

/transl_table=11

/product="4-hydroxybenzoate synthetase"

/protein_id="CAA40681.1"

/db_xref="GI:43234"

/db_xref="GOA:P26602"

/db_xref="UniProtKB/Swiss-Prot:P26602"

/translation="MRLRFCCVLDHLICFTSPVNTFLRYNAFTLCNGFCGMSHPALT QRLARYCKEIPALDPQLDLWELLEDSTKRFQGGKTVSVTMIREFGVEQNEIPEEL PLLPKESYWLREILLCADGEPFLAGRTVPVPSGLPELALQKLGKTPPLGRYLFTSS TLRDFIETGRDAGLWGRRLRLSGKPLLTFLFASPPLY"

gene 912..1784

/gene="ubiA"

CDS 912..1784

/gene="ubiA"

/codon_start=1

/transl_table=11

/product="4-hydroxybenzoate octaprenyltransferase"

/protein_id="CAA40682.1"

/db_xref="GI:43235"

/db_xref="GOA:P26601"

/db_xref="UniProtKB/Swiss-Prot:P26601"

/translation="MWSLTQNLKLAFLHRLMRTDKPIGALLLWPTLMALWVATPGVP QLWILAVFVAGVWLMRAAGCVVNDYADRKFDGHVKRTANRPLPSGAVTEKEARALFVV LVLISFLVLTNTMTILSLTAALAWVPFMKRYTHLPQVILGAAGFWSIPMAFAA VSESVPLSCWLMFLANILMAVAYDTQYAWDRDDDKIGIKSTAILFGQYDKLIIGIL QIVGLMALWITIGELNGWGIYSILVAGALFVYQQKLIANREREACFKAFMNNYVG LVLFLGLAMSYWHF"

ORIGIN

Alignment Scores:

Pred. No.:	7,15e-81	Length:	2034
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	15	Gaps:	0

US-10-718-311-4 (1-165) x ECUBIAC (1-2034)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
 |||||
 Db 402 ATGTACACCCCGCGCTTAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 461
 |||||

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
 |||||
 Db 462 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGGAA 521
 |||||

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
 |||||
 Db 522 CAGCAGGGGAAAACCGTAAGCGTGACGATGATCCCGAAGGGGTTTGTCCAGCAGAAATGAA 581
 |||||

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
 |||||
 Db 582 ATCCCGGAAGAACTGCCCGCTGTCGCCGAAGAGATCTCGTTACTCGTTACGTGAAATTTG 641
 |||||

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
 |||||
 Db 642 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTACCGTCTCTGTGTCAACGTTA 701
 |||||

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
 |||||
 Db 702 AGCGGGCGGAGCTGGCGTTACAAAATTTGGGTAAACGCCGTTAGGACGCTATCTGTTT 761
 |||||

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
 |||||
 Db 762 ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCCGCTGATGCCGGGCTGTGGGG 821
 |||||

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
 |||||
 Db 822 CGACGTTCCCGCCCTGCGATTAAAGCGGTAAACCGCTGTGTGCTAACAGAACTGTTTTACCG 881
 |||||

Qy 161 AlaSerProLeuTyr 165
 |||||

Db 882 GCGTCACCGTTGTATC 896
 |||||

Search completed: May 30, 2006, 05:53:01
 Job time : 5004.73 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2006, 11:01:13 ; Search time 200.955 Seconds
(without alignments)
2304.491 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALTOALRYLCKEIPA.....RLSGKPLLLTELFLPASPLY 165

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1403666 seqs, 93554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_p2n.model -DEV=xlp
-O=/abss/ABSSWEB spool/US10718311/runat 26052006_164922_13534/app_query.fasta_1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs02p -USER=US10718311 @CGN 1.1 615 @runat 26052006_164922_13534
-NCPU=6 -ICPU=3 -NO MAP -NEG SCORES=0 -WAIT -DSBLOK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /EMC_Celerra_SIDS3/ptodata/2/ina/1 COMB.seq.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/ina/5 COMB.seq.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/ina/6A COMB.seq.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/ina/6B COMB.seq.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/ina/7 COMB.seq.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/ina/H COMB.seq.*
- 7: /EMC_Celerra_SIDS3/ptodata/2/ina/PCUTUS COMB.seq.*
- 8: /EMC_Celerra_SIDS3/ptodata/2/ina/PP COMB.seq.*
- 9: /EMC_Celerra_SIDS3/ptodata/2/ina/RE COMB.seq.*
- 10: /EMC_Celerra_SIDS3/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	854	100.0	495	3	US-09-896-866B-12 Sequence 12, Appl
2	854	100.0	495	3	US-09-855-341-3 Sequence 3, Appl
3	854	100.0	684	3	US-09-896-866B-16 Sequence 16, Appl
4	854	100.0	684	3	US-09-855-341-7 Sequence 7, Appl
5	854	100.0	854	3	US-09-064-693A-24 Sequence 24, Appl
6	854	100.0	6641	3	US-09-064-693A-25 Sequence 25, Appl
7	673	78.8	645	3	US-09-489-039A-1234 Sequence 1234, Ap
8	401	47.0	537	3	US-09-543-681A-3122 Sequence 3122, Ap

9	186	21.8	630	3	US-09-252-991A-7442	Sequence 7442, Ap
10	174.5	20.4	519	3	US-09-540-236-1889	Sequence 1889, Ap
11	171.5	20.1	269223	3	US-09-596-002-41	Sequence 41, Appl
12	157.5	18.4	540	3	US-09-328-352-343	Sequence 343, Appl
13	146.5	17.2	1248	3	US-09-252-991A-7361	Sequence 7361, Ap
14	121	14.2	195	3	US-09-543-681A-3086	Sequence 3086, Ap
15	90	10.5	2127	3	US-09-818-780-75	Sequence 75, Appl
16	83	9.7	84839	3	US-09-949-016-15816	Sequence 15816, A
17	82	9.6	601	3	US-09-949-016-69693	Sequence 69693, A
18	82	9.6	807	3	US-09-583-110-1232	Sequence 1232, Ap
19	82	9.6	888	3	US-09-107-433-1721	Sequence 1721, Ap
20	81	9.5	62354	3	US-09-949-016-16188	Sequence 16188, A
21	80	9.4	6463	2	US-08-962-284-3	Sequence 3, Appli
22	80	9.4	68778	3	US-09-949-016-16406	Sequence 16406, A
23	80	9.4	72455	3	US-09-949-016-13793	Sequence 13793, A
24	80	9.4	77535	3	US-09-949-016-14279	Sequence 14279, A
25	80	9.4	77535	3	US-09-949-016-14280	Sequence 14280, A
26	80	9.4	77535	3	US-09-949-016-14281	Sequence 14281, A
27	80	9.4	104475	3	US-09-949-016-12115	Sequence 12115, A
28	80	9.4	111282	3	US-09-754-250-3	Sequence 3, Appli
29	80	9.4	111282	3	US-10-094-989-3	Sequence 3, Appli
30	79.5	9.3	47981	3	US-09-679-279-1	Sequence 1, Appli
31	79.5	9.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
32	79.5	9.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
33	79	9.3	1176	3	US-09-252-991A-10640	Sequence 10640, A
34	79	9.3	1581	3	US-09-252-991A-10731	Sequence 10731, A
35	79	9.3	2055	3	US-09-252-991A-10433	Sequence 10433, A
36	79	9.3	5246	3	US-09-799-451-474	Sequence 474, App
37	78.5	9.2	9461	3	US-09-221-017B-513	Sequence 513, App
38	78	9.1	4792	3	US-08-781-891-205	Sequence 205, App
39	78	9.1	4792	3	US-09-618-166-205	Sequence 205, App
40	78	9.1	9173	3	US-09-949-001-30	Sequence 30, Appl
41	78	9.1	9174	2	US-09-949-001-36	Sequence 36, Appl
42	78	9.1	49377	2	US-08-764-233A-1	Sequence 1, Appli
43	77.5	9.1	1090	3	US-09-533-559-3860	Sequence 3860, Ap
44	77.5	9.1	57392	3	US-09-949-016-12070	Sequence 12070, A
45	77.5	9.1	57402	3	US-09-949-016-13293	Sequence 13293, A

ALIGNMENTS

RESULT 1

US-09-896-866B-12
; Sequence 12, Application US/09896866B
; Patent No. 6642036
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12

Alignment Scores:			
Pred. No.:	5.04e-98	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	3	Gaps:	0
US-10-718-311-4 (1-165) x US-09-896-866B-12 (1-495)			

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 1 ATGTACACCCCGCGCTTAACCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60

Qy 21 LeuAepProGlnLeuAepTrpLeuLeuLeuGluAepSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTGCATGCGTGTGTTGGAGGATTTCATGACAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGCAAAACGGTAAGCGTGACGATATCCGCAAGGCTTGTTCGACGAGATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 181 ATCCCGCAAGAACTGCGCTGTCGCAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240

Qy 81 LeuCysAlaAepGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCGATGTTGAACCGTGGCTTCCCGTGGTACCGTCTCTGTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGGAGCTGGCGTTACAAAATTGGTAAACCGCTTAGGACGCTATCTGTTC 360

Qy 121 ThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGTTCGCCCTGGATTAAAGCGGTAAACCGCTGTGCTTAACGAACTGTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 CGGTACCGTGTATAC 495

RESULT 2

US-09-855-341-3
; Sequence 3, Application US/09855341
; Patent No. 6683231

; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNU
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: IN GREEN PLANTS
; CURRENT APPLICATION NUMBER: US/09/855,341
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli

US-09-855-341-3

Alignment Scores:
Pred. No.: 5,04e-98 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-855-341-3 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 1 ATGTACACCCCGCGCTTAACCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60

Qy 21 LeuAepProGlnLeuLeuAepTrpLeuLeuLeuGluAepSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTGCATGCGTGTGTTGGAGGATTTCATGACAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGCAAAACGGTAAGCGTGACGATATCCGCAAGGCTTGTTCGACGAGATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 181 ATCCCGCAAGAACTGCGCTGTCGCAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240

Qy 81 LeuCysAlaAepGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCGATGTTGAACCGTGGCTTCCCGTGGTACCGTCTCTGTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGGAGCTGGCGTTACAAAATTGGTAAACCGCTTAGGACGCTATCTGTTC 360

Qy 121 ThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGTTCGCCCTGGATTAAAGCGGTAAACCGCTGTGCTTAACGAACTGTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 CGGTACCGTGTATAC 495

RESULT 3

US-09-896-866B-16
; Sequence 16, Application US/09896866B
; Patent No. 6642036

; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylgucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein

US-09-896-866B-16

Alignment Scores:
Pred. No.: 8,29e-98 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-896-866B-16 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 187 ATGTACACCCCGCGCTTAACGCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 246

QY 21 LeuAspProGluLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTGA 306
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGGAAAAACGGTAGCGTACGATGATCCGGAAGGGTTCGAGCAGAAATGAA 366
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTAGTGAATTTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGTTCAACGTTA 486
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTrpLeuPhe 120
Db 487 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTATTAGATAGGCGGTGATGCCGGCTGTGGGG 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGACGTTCCCGCTCGATTAAAGCGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 4

US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. 668231
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: IN GREEN PLANTS
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7

Alignment Scores:
Pred. No.: 8,298-98 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-855-341-7 (1-684)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTrpCysGlyGluIleProLa 20
Db 187 ATGTCACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGAGATCCCTGCC 246
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGTTTGA 306

QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGGAAAAACGGTAGCGTACGATGATCCGGAAGGGTTCGAGCAGAAATGAA 366
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTAGTGAATTTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGTTCAACGTTA 486
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTrpLeuPhe 120
Db 487 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTATTAGATAGGCGGTGATGCCGGCTGTGGGG 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGACGTTCCCGCTCGATTAAAGCGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 5

US-09-064-693A-24
; Sequence 24, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INBEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-24

Alignment Scores:
Pred. No.: 1,17e-97 Length: 854
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-064-693A-24 (1-854)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db |||||
Qy 321 ATGTCACACCCCGGTTAACGCACTGCTGCGCTGCGCTATTGTAAGAGATCCCTGCC 380
Db |||||
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db |||||
Qy 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACACAAACGTTTGA 440
Db |||||
Qy 41 GlnGlnGlyLysThrValSerValThrMetileArgGluGlyPheValGluGlnAsnGlu 60
Db |||||
Qy 441 CAGCAGGGAAACCGTAAAGCGTACGATGATCCGCGAAGGGTTTGTGAGCAGAAATGAA 500
Db |||||
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileu 80
Db |||||
Qy 501 ATCCCGGAAGAACTGCCCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 560
Db |||||
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db |||||
Qy 561 TTATGTCCGATGGTGAACCGTGGCTTCCGCGTCTGATCCGTCCTGTCACAGTTA 620
Db |||||
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrleuPhe 120
Db |||||
Qy 621 AGCGGGCCGAGCTGGCGTTACAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 680
Db |||||
Qy 121 ThrSerSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrpGly 140
Db |||||
Qy 681 ACATCATCGACATTAACCCGGAGCTTATTAGATAGGCCGTGATGCCGGCTGTGGGG 740
Db |||||
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuLeuLeuLeuLeuPro 160
Db |||||
Qy 741 CGAGCTTCCCGCTCGGATTAAGCGGTAACCGCTTGTGCTTAACAGAACTGTTTTTACC 800
Db |||||
Qy 161 AlaSerProLeuTyr 165
Db |||||
Qy 801 GCGTCACCGTTGTAC 815

RESULT 6
US-09-064-693A-25
; Sequence 25, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESS: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208) 526-8339
; TELEFAX: (208) 526-8339
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-25

Alignment Scores:
Pred. No.: 2,74e-96 Length: 6641
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-064-693A-25 (1-6641)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db 321 ATGTCACACCCCGGTTAACGCACTGCTGCGCTGCGCTATTGTAAGAGATCCCTGCC 380
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACACAAACGTTTGA 440
Qy 41 GlnGlnGlyLysThrValSerValThrMetileArgGluGlyPheValGluGlnAsnGlu 60
Db 441 CAGCAGGGAAACCGTAAAGCGTACGATGATCCGCGAAGGGTTTGTGAGCAGAAATGAA 500
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileu 80
Db 501 ATCCCGGAAGAACTGCCCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 560
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 561 TTATGTCCGATGGTGAACCGTGGCTTCCGCGTCTGATCCGTCCTGTCACAGTTA 620
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrleuPhe 120
Db 621 AGCGGGCCGAGCTGGCGTTACAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 680
Qy 121 ThrSerSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrpGly 140
Db 681 ACATCATCGACATTAACCCGGAGCTTATTAGATAGGCCGTGATGCCGGCTGTGGGG 740
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuLeuLeuLeuLeuPro 160
Db 741 CGAGCTTCCCGCTCGGATTAAGCGGTAACCGCTTGTGCTTAACAGAACTGTTTTTACC 800
Qy 161 AlaSerProLeuTyr 165
Db 801 GCGTCACCGTTGTAC 815

RESULT 7
US-09-489-039A-1234
; Sequence 1234, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 2709.2004001
;; CURRENT APPLICATION NUMBER: US/09/489,039A
;; CURRENT FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: US 60/117,747
;; PRIOR FILING DATE: 1999-01-29
;; NUMBER OF SEQ ID NOS: 14342
;; SEQ ID NO 1234
;; LENGTH: 645
;; TYPE: DNA
;; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1234

Alignment Scores:
Pred. No.: 4,54e-75 Length: 645
Score: 673.00 Matches: 129
Percent Similarity: 87.9% Conservatives: 16
Best Local Similarity: 78.2% Mismatches: 20
Query Match: 78.8% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-489-039A-1234 (1-645)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 148 ATGTCCCATCTCGCGCTTACGGGACTGCGTGGCGCTATTTTGGCGGTATGCCATCC 207
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 208 CTGCGCGCCCGCTAGTCACTGGCTGCTGCTGGAGGACTCAATGACCCAGCGCTTGAA 267
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 268 CAACAAGGAACAGGTACCGTACCGTGGTTACGAAGGTATATCGCGCGTACGGG 327
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 328 CTGACTGACGAAGCGCGCTGCTGCCGACGAGCGCGCTACTGCGTGGCGGAGATTATA 387
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 388 CTCAATGCGAGTGGCGAGCCCTGGCTGGCGGCGCACGCTGGCGCGAGTCGACGCTG 447
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 448 TGTGTCGGAGCTGGCGCTACAGAGCTCGGGCAACTCCGCTGGCGGATACCTGTTT 507
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140
Db 508 ACGTCGTCGACATTAAACCGCGATTATTATTGAAATTGGTTCGCGATCGACGCGCTGTGGGG 567
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 568 CGTCGTCCTCCGCTGCGGTGAGCGGCAAAACCCCTGCTGCTACCGAGCTTTTGTGCT 627
QY 161 AlaSerProLeuTyr 165
Db 628 GGTGCGCGTGTAC 642

RESULT 8
US-09-543-681A-3122
; Sequence 3122, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344

;; SEQ ID NO 3122
;; LENGTH: 537
;; TYPE: DNA
;; ORGANISM: Proteus mirabilis
US-09-543-681A-3122

Alignment Scores:
Pred. No.: 5.8e-41 Length: 537
Score: 401.00 Matches: 81
Percent Similarity: 68.6% Conservatives: 15
Best Local Similarity: 57.9% Mismatches: 44
Query Match: 47.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-543-681A-3122 (1-537)

QY 26 LeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThr 45
Db 109 TTAAGTTGGCTGCTGGAATTAGGTTCATATGACACGACGTTTTCAGCAACATTGCCATCA 168
QY 46 ValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluLeuProGluGluLeu 65
Db 169 GTGACGGTAATGCCATATCAAGAAGTTTATTGAATATATATCGAGCTCTGCTGATGAACA 228
QY 66 ProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAspGly 85
Db 229 AGTGTTTTACCTTATAGCCGACGCTATTGGCTTAGAAGAAATGTCCTTTTGTGGGATAAT 288
QY 86 GluProTrpLeuAlaGlyArgThrValValProValSerThrLeuSerGlyProGluLeu 105
Db 289 GTACCTTGGTTATTAGGCGCAACATTAGTCCGGAAGACATTACCGGTGAAGATCGC 348
QY 106 AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerSerThrLeu 125
Db 349 CAATTAGTGAATTGGTACGTCAGCGCTATTAGGACGTTATCTGTTCAGGAACAACCTTA 408
QY 126 ThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeu 145
Db 409 AGTCGTGATTATTATTCATATTGGGCAACAAATGGACATTGGTACGTCTCCGCTTC 468
QY 146 ArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165
Db 469 CAGCTTTCAGATAAACCTTTATTATTACTAGAGGTGTTTTTGCCTGCATCACCTGTATAT 528

RESULT 9
US-09-252-991A-7442
; Sequence 7442, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7442
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7442

Alignment Scores:
Pred. No.: 8.44e-14 Length: 630
Score: 186.00 Matches: 50
Percent Similarity: 43.0% Conservatives: 21
Best Local Similarity: 30.3% Mismatches: 64
Query Match: 21.8% Indels: 30
DB: 3 Gaps: 3

RESULT 13
US-09-252-991A-7361
; Sequence 7361, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOTUBERCULOSIS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

Pred. No.:	2.26e-08	Length:	1248
Score:	146.50	Matches:	35
Percent Similarity:	46.4%	Conservative:	10
Best Local Similarity:	36.1%	Mismatches:	41
Query Match:	17.2%	Indels:	11
DB:	3	Gaps:	1
US-10-718-311-4 (1-165)	x	US-09-253-991A-7361 (1-1248)	

DB 65 GCGGCGCGCAGCGCCCTGGGAAGGCTCGGGCTTCGACCTGGCGCTGCTCGGCACCCGCTCG 124

RESULT 14
US-09-543-681A-3086/C
; Sequence 3086, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEIN
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - nucleic search, using frame_plus_p2n model
Run on: May 30, 2006, 03:03:17 ; Search time 1135.79 Seconds
(without alignments)
2677.596 Million cell updates/sec

Title: US-10-718-311-4
Perfect score: 854
Sequence: 1 MSHPALQRLALRYCKEIPA.....RLSGKPLLTLELPASPLY 165

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/abs/ABSWEB/spool/US10718311/runat_26052006_164933_13735/app_query.fasta_1
-DB=PublishedApplications_NA_Main -QFMT=fastap -SUFFIX=rnpbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abs805h
-USER=US10718311 @CGN 1.1.2326 @runat_26052006_164933_13735 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : Published Applications NA_Main:
1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq.*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10J_PUBCOMB.seq.*
15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10K_PUBCOMB.seq.*
16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10L_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	854	100.0	495	3	US-09-855-341-3 Sequence 3, Appli

2	854	100.0	495	3	US-09-896-866B-12	Sequence 12, Appl
3	854	100.0	495	7	US-10-359-369-37	Sequence 37, Appl
4	854	100.0	495	7	US-10-699-050-12	Sequence 12, Appl
5	854	100.0	495	8	US-10-718-311-3	Sequence 3, Appli
6	854	100.0	498	3	US-10-462-162-29	Sequence 29, Appl
7	854	100.0	684	3	US-09-855-341-7	Sequence 7, Appli
8	854	100.0	684	3	US-09-896-866B-16	Sequence 16, Appl
9	854	100.0	684	7	US-10-359-369-41	Sequence 41, Appl
10	854	100.0	684	8	US-10-699-050-16	Sequence 16, Appl
11	854	100.0	684	8	US-10-718-311-7	Sequence 7, Appli
12	854	100.0	684	9	US-10-462-162-33	Sequence 33, Appl
13	854	100.0	3452	9	US-10-462-162-47	Sequence 47, Appl
14	854	98.7	1971	9	US-10-462-162-54	Sequence 54, Appl
15	811	95.0	1207	10	US-10-450-763-26139	Sequence 26139, A
16	171.5	20.1	269233	8	US-10-672-787-41	Sequence 41, Appl
17	97.5	11.4	475	8	US-10-275-026A-113	Sequence 113, App
18	91	10.7	606	4	US-09-925-065A-789463	Sequence 789463,
19	91	10.7	606	5	US-09-925-065A-789463	Sequence 789463,
20	90	10.5	2127	9	US-10-746-167-75	Sequence 75, Appl
21	90	10.5	9133	10	US-10-915-740A-277	Sequence 277, App
22	90	10.5	2242716	10	US-10-915-740A-1068	Sequence 1068, Ap
23	88	10.3	42811	11	US-10-330-773-925	Sequence 925, App
24	87	10.2	59589	6	US-10-087-192-1315	Sequence 1315, Ap
25	86	10.1	2892	16	US-11-183-136-15	Sequence 15, Appl
26	85.5	10.0	522	10	US-10-467-657-2855	Sequence 2855, Ap
27	84.5	9.9	1780	10	US-10-487-078-92	Sequence 92, Appl
28	84.5	9.9	111331	10	US-10-461-862-101	Sequence 101, App
29	83.5	9.8	1565	10	US-10-750-185-33285	Sequence 33285, A
30	83.5	9.8	1565	10	US-10-750-623-33285	Sequence 33285, A
31	83.5	9.8	10710	8	US-10-672-396-4	Sequence 4, Appli
32	83	9.7	608	4	US-09-925-065A-772942	Sequence 772942,
33	83	9.7	608	5	US-09-925-065A-772942	Sequence 772942,
34	82.5	9.7	689	13	US-11-097-143-18275	Sequence 18275, A
35	82.5	9.7	122336	8	US-10-322-281-362	Sequence 362, App
36	82	9.6	444	12	US-10-301-480-289318	Sequence 289318,
37	82	9.6	444	12	US-10-301-480-902727	Sequence 902727,
38	82	9.6	452	4	US-09-925-065A-200730	Sequence 200730,
39	82	9.6	452	5	US-09-925-065A-200730	Sequence 200730,
40	82	9.6	888	10	US-10-617-320-1721	Sequence 1721, Ap
41	81.5	9.5	6369	8	US-10-437-963-11402	Sequence 11402, A
42	81	9.5	435	4	US-09-925-065A-200729	Sequence 200729,
43	81	9.5	435	5	US-09-925-065A-200729	Sequence 200729,
44	81	9.5	439	12	US-10-301-480-289317	Sequence 289317,
45	81	9.5	439	12	US-10-301-480-902726	Sequence 902726,

ALIGNMENTS

RESULT 1
US-09-855-341-3
; Sequence 3, Application US/09855341
; Patent No. US2002002715A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-09-855-341-3

Alignment Scores:
Pred. No.: 1.91e-102 Length: 495

Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-855-341-3 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 1 ATGTACACACCCCGCGTTAACCAACTCGCTGGCTGGCTATTTGTAAGAGATCCCTGCC 60

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60
Db 121 CAGCAGGGAACCGTAAGCGTGACGATGATCCGGAAGGGTTTGTGACAGAGATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCGAGTGGTAACCGTGGCTGCTCCCGTACCGTCTGTTACGTGAAATTTG 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGGCGGAGTGGCTGTTACNAAAATTTGGTAAACCGCTTAGGACGCTATCTGTT 360

Qy 121 ThrSerThrLeuThrArgAspPheIleGluLeGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGCGTGTGCGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGACGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTATC 495

RESULT 2
US-09-896-866B-12
; Sequence 12, Application US/09896866B
; Patent No. US20020151002A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylgucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR FILING DATE: 60/216,615
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12

Alignment Scores:
Pred. No.: 1.91e-102 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0

DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-896-866B-12 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 1 ATGTACACACCCCGCGTTAACCAACTCGCTGGCTGGCTATTTGTAAGAGATCCCTGCC 60

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60
Db 121 CAGCAGGGAACCGTAAGCGTGACGATGATCCGGAAGGGTTTGTGACAGAGATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCGAGTGGTAACCGTGGCTGCTCCCGTACCGTCTGTTACGTGAAATTTG 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGGCGGAGTGGCTGTTACAAAAATTTGGTAAACCGCTTAGGACGCTATCTGTT 360

Qy 121 ThrSerThrLeuThrArgAspPheIleGluLeGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGCGTGTGCGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGACGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTATC 495

RESULT 3
US-10-359-369-37
; Sequence 37, Application US/10359369
; Publication No. US20030215927A1
; GENERAL INFORMATION:
; APPLICANT: E. I. duPont de Nemours and Company, Inc.
; APPLICANT: Viitanen, Paul
; APPLICANT: Meyer, Knut
; APPLICANT: Van Dyk, Drew
; TITLE OF INVENTION: UDP-Glucosyltransferases
; FILE REFERENCE: CUI821 US NA
; CURRENT APPLICATION NUMBER: US/10/359,369
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: 60/355,511
; PRIOR FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(495)
US-10-359-369-37

Alignment Scores:
Pred. No.: 1.91e-102 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0

US-10-718-311-4 (1-165) x US-10-359-369-37 (1-495)

```

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|
|
|
Db 1 ATGTACACCCCGGTTAACGCAACTGCGTGGCTATTTGTAAGAGATCCCTGCC 60
|
|
|
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|
|
|
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTTGAA 120
|
|
|
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnGlnAenGlu 60
|
|
|
Db 121 CAGCAGGGGAAAAACGGTAAGCGTGACGATCCGCGAAGGGTTTGTCTGAGCAGAAATGAA 180
|
|
|
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 181 ATCCCCGGAAGACTGCGCGTGTCCGCAAGAGTCTCGTTACTGTTAGTGAATTTTG 240
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
|
|
|
Db 241 TTATGTGCGATGGTGAACCGTGGCTTGCCTGCTGCTGTTAGTGAATTTTG 300
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 301 AGCGGGCCGAGCTGCGGTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
|
|
|
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 361 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTAGTCCGGCTGTGGGG 420
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 421 CGACGTTCCCGCTCGATTAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 480
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 481 GCGTCACCGTTGTAC 495

```

RESULT 4

US-10-699-050-12

; Sequence 12, Application US/10699050

; Publication NO. US20040142437A1

; GENERAL INFORMATION:

; APPLICANT: Meyer, Knut

; APPLICANT: Flint, Dennis

; APPLICANT: Viitanen, Paul

; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates

; FILE REFERENCE: Benzoic Acid Glucosides

; CURRENT APPLICATION NUMBER: US/10/699,050

; PRIOR FILING DATE: 2003-10-30

; PRIOR APPLICATION NUMBER: US/09/896,866B

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: 60/216,615

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 12

; LENGTH: 495

; TYPE: DNA

; ORGANISM: Escherichia coli

US-10-699-050-12

Alignment Scores:

Pred. No.:	1,918-102	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	8	Gaps:	0

US-10-718-311-4 (1-165) x US-10-699-050-12 (1-495)

```

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|
|
|
Db 1 ATGTACACCCCGGTTAACGCAACTGCGTGGCTATTTGTAAGAGATCCCTGCC 60
|
|
|
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|
|
|
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTTGAA 120
|
|
|
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnGlnAenGlu 60
|
|
|
Db 121 CAGCAGGGGAAAAACGGTAAGCGTGACGATCCGCGAAGGGTTTGTCTGAGCAGAAATGAA 180
|
|
|
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 181 ATCCCCGGAAGACTGCGCGTGTCCGCAAGAGTCTCGTTACTGTTAGTGAATTTTG 240
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
|
|
|
Db 241 TTATGTGCGATGGTGAACCGTGGCTTGCCTGCTGCTGTTAGTGAATTTTG 300
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 301 AGCGGGCCGAGCTGCGGTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
|
|
|
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 361 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTAGTCCGGCTGTGGGG 420
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 421 CGACGTTCCCGCTCGATTAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 480
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 481 GCGTCACCGTTGTAC 495

```

RESULT 5

US-10-718-311-3

; Sequence 3, Application US/10718311

; Publication NO. US20040143867A1

; GENERAL INFORMATION:

; APPLICANT: MEYER, KNU

; APPLICANT: VIITANEN, PAUL V.

; APPLICANT: VAN DYK, DREW

; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID

; TITLE OF INVENTION: IN GREEN PLANTS

; FILE REFERENCE: BC1015 US NA

; CURRENT APPLICATION NUMBER: US/10/718,311

; PRIOR FILING DATE: 2003-11-20

; PRIOR APPLICATION NUMBER: US/09/855,341

; PRIOR FILING DATE: 2001-05-15

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: MICROSOFT OFFICE 97

; SEQ ID NO 3

; LENGTH: 495

; TYPE: DNA

; ORGANISM: Unknown Organism

; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism:E. coli

US-10-718-311-3

Alignment Scores:

Pred. No.:	1,918-102	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	8	Gaps:	0

US-10-718-311-4 (1-165) x US-10-718-311-3 (1-495)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20

```
Db 1 ATGTCACACCCCGGTTAAACCACTGCTGGCTGCCTATTGTAAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATCCATGACAAAACGTTTGAA 120
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGGAAAAACGGTAAGCGTGACGATATCCGCGAAGGGTTTGTGACGAGAATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTGCCGGTGGTACCGTCTTCTGTGCAACGTTA 300
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGGCGGAGCTGGCGGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTT 360
Qy 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495
RESULT 6
US-10-462-162-29
; Sequence 29, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-29
Alignment Scores:
Pred. No.: 1.93e-102 Length: 498
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0
US-10-718-311-4 (1-165) x US-10-462-162-29 (1-498)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGGTTAAACCACTGCTGGCTGCCTATTGTAAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATCCATGACAAAACGTTTGAA 120
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGGAAAAACGGTAAGCGTGACGATATCCGCGAAGGGTTTGTGACGAGAATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTGCCGGTGGTACCGTCTTCTGTGCAACGTTA 300
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGGCGGAGCTGGCGGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTT 360
Qy 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495
```

```
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGGAAAAACGGTAAGCGTGACGATATCCGCGAAGGGTTTGTGACGAGAATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTGCCGGTGGTACCGTCTTCTGTGCAACGTTA 300
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGGCGGAGCTGGCGGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTT 360
Qy 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495
RESULT 7
US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. US20020002715A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7
Alignment Scores:
Pred. No.: 3e-102 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0
US-10-718-311-4 (1-165) x US-09-855-341-7 (1-684)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTCACACCCCGGTTAAACCACTGCTGGCTGCCTATTGTAAAGAGATCCCTGCC 246
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATCCATGACAAAACGTTTGAA 306
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGGAAAAACGGTAAGCGTGACGATATCCGCGAAGGGTTTGTGACGAGAATGAA 366
```

```
QY 61 IleProGluLeuProLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 367 ATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGGTTACGTGAATTTTG 426
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
|
|
|
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTCTGTCAACGTTA 486
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
|
|
|
QY 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 547 ACATCATCGACATTAACCGGACCTTTATTGAGTAGGCGCGTATCGCGGCTGTGGGG 606
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 607 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTCGTAAACAGAACTGTTTTTACCG 666
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 667 GCGTCACCGTTGTAC 681
|
|
|
RESULT 8
US-09-896-866B-16
; Sequence 16, Application US/09896866B
; Patent No. US20020151002A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose-Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein
US-09-896-866B-16
Alignment Scores:
Pred. No.: 3e-102 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0
US-10-718-311-4 (1-165) x US-09-896-866B-16 (1-684)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|
|
|
Db 187 ATGTACACACCGCGTTAACGCAACTGCGTGGCTGGCTATTGTAAAGAGATCCCTGCC 246
|
|
|
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|
|
|
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTTCGTGGAGATTCCATGACAAAACGTTTGA 306
|
|
|
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
|
|
|
Db 307 CAGCAGGGGAAAACCGTAGCGTAGATGATCCCGAAGGGTGTTCGAGCAGAAATGAA 366
|
|
|
QY 61 IleProGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 367 ATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGGTTACGTGAATTTTG 426
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
|
|
|
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTCTGTCAACGTTA 486
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
|
|
|
QY 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 547 ACATCATCGACATTAACCGGACCTTTATTGAGTAGGCGCGTATCGCGGCTGTGGGG 606
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 607 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTCGTAAACAGAACTGTTTTTACCG 666
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 667 GCGTCACCGTTGTAC 681
|
|
|
RESULT 9
US-10-359-369-41
; Sequence 41, Application US/10359369
; Publication No. US20030215927A1
; GENERAL INFORMATION:
; APPLICANT: E. I. duPont de Nemours and Company, Inc.
; APPLICANT: Viitanen, Paul
; APPLICANT: Meyer, Knut
; APPLICANT: Van Dyk, Drew
; TITLE OF INVENTION: UDP-Glucosyltransferases
; FILE REFERENCE: C11821 US NA
; CURRENT APPLICATION NUMBER: US/10/359,369
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: 60/355,511
; PRIOR FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 41
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(684)
US-10-359-369-41
Alignment Scores:
Pred. No.: 3e-102 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0
US-10-718-311-4 (1-165) x US-10-359-369-41 (1-684)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|
|
|
Db 187 ATGTACACACCGCGTTAACGCAACTGCGTGGCTGGCTATTGTAAAGAGATCCCTGCC 246
|
|
|
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|
|
|
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTTCGTGGAGATTCCATGACAAAACGTTTGA 306
|
|
|
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
|
|
|
Db 307 CAGCAGGGGAAAACCGTAGCGTAGATGATCCCGAAGGGTGTTCGAGCAGAAATGAA 366
|
|
|
QY 61 IleProGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 367 ATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGGTTACGTGAATTTTG 426
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
|
|
|
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTCTGTCAACGTTA 486
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
|
|
|
QY 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 547 ACATCATCGACATTAACCGGACCTTTATTGAGTAGGCGCGTATCGCGGCTGTGGGG 606
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 607 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTCGTAAACAGAACTGTTTTTACCG 666
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 667 GCGTCACCGTTGTAC 681
|
|
|
```

```

Db      367 ATCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAATTTTG 426
Qy      81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db      427 TTATGTGCGGATGGTGAACCGTGGCTTCCCGGTGCTACCGTCTCTCTGTCAACGTTA 486
Qy     101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrlieuPhe 120
Db      487 AGCGGGCGGAGCTGGCGGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTC 546
Qy     121 ThrSerSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrpGly 140
Db      547 ACATCATCGACATTAACCCGGGACTTATTAGATAGGCCGCTGATGCCGGCTGTGGGG 606
Qy     141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db      607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
Qy     161 AlaSerProLeuTyr 165
Db      667 GCGTCACCGTTGTAC 681

```

RESULT 10

```

US-10-699-050-16
; Sequence 16, Application US/10699050
; Publication No. US20040142437A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; TITLE OF INVENTION: Benzoic Acid Glucosides
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/10/699,050
; CURRENT FILING DATE: 2003-10-30
; PRIOR APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein
US-10-699-050-16

```

```

Alignment Scores:
Pred. No.:      3e-102      Length:      684
Score:          854.00      Matches:      165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:      0
DB:              8          Gaps:          0

```

US-10-718-311-4 (1-165) x US-10-699-050-16 (1-684)

```

Qy      1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db     187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCTATTGTTAAAGAGATCCCTGCC 246
Qy     21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db     247 CTGGATCCGCAACTGCTCGACTGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306
Qy     41 GlnGlnGlyLysThrValSerValThrMetileArgGluGlyPheValGluGlnAsnGlu 60
Db     307 CAGCAGGAGAAAACCGGTAAAGCGTGACGATGATCCCGAAGGGTTTGTTCGAGCAGATGAA 366

```

```

Qy     61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80
Db     367 ATCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAATTTTG 426
Qy     81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db     427 TTATGTGCGGATGGTGAACCGTGGCTTCCCGGTGCTACCGTCTCTCTGTCAACGTTA 486
Qy     101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrlieuPhe 120
Db     487 AGCGGGCGGAGCTGGCGGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTC 546
Qy     121 ThrSerSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrpGly 140
Db     547 ACATCATCGACATTAACCCGGGACTTATTAGATAGGCCGCTGATGCCGGCTGTGGGG 606
Qy     141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db     607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
Qy     161 AlaSerProLeuTyr 165
Db     667 GCGTCACCGTTGTAC 681

```

RESULT 11

```

US-10-718-311-7
; Sequence 7, Application US/10718311
; Publication No. US20040143867A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/10/718,311
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: US/09/855,341
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; NAME/KEY: misc feature
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-10-718-311-7

```

```

Alignment Scores:
Pred. No.:      3e-102      Length:      684
Score:          854.00      Matches:      165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:      0
DB:              8          Gaps:          0

```

US-10-718-311-4 (1-165) x US-10-718-311-7 (1-684)

```

Qy      1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db     187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCTATTGTTAAAGAGATCCCTGCC 246
Qy     21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db     247 CTGGATCCGCAACTGCTCGACTGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306
Qy     41 GlnGlnGlyLysThrValSerValThrMetileArgGluGlyPheValGluGlnAsnGlu 60
Db     307 CAGCAGGAGAAAACCGGTAAAGCGTGACGATGATCCCGAAGGGTTTGTTCGAGCAGATGAA 366
Qy     61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80

```

Db 367 ATCCCCGAAGAACTGCCGCTGTCGCCGAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCTGACCGTCTGTTCTGTCAACGTTA 486
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 487 AGCGGGCCGGAGCTGGCGTTACAAAATTTGGGTAAACCGCCGTTAGGACGCTATCTGTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTGATCCGGGCTGTGGGG 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGACGTTCCCGCTGGATTAAAGCGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 12

US-10-462-162-33
; Sequence 33, Application US/10462162
; Publication NO. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-33

Alignment Scores:
Pred. No.: 3e-102 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-10-718-311-4 (1-165) x US-10-462-162-33 (1-684)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTACACCCCGCTTAACGCAACTGCGTGGCTATTTAAAGAGATCCCTGCC 246
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTGACGTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 306
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGGAAAAACGGTAAGCTGACGATGATCCCGAAGGTTGTGCGAGCAGATGAA 366
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCCGAAGAACTGCCGCTGTCGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGACCGTCTGTTCTGTCAACGTTA 486

QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 487 AGCGGGCCGGAGCTGGCGTTACAAAATTTGGGTAAACCGCCGTTAGGACGCTATCTGTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTGATCCGGGCTGTGGGG 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGACGTTCCCGCTGGATTAAAGCGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 13

US-10-462-162-47
; Sequence 47, Application US/10462162
; Publication NO. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 47
; LENGTH: 3452
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of nucleic acid fragment encoding CPL (from
; OTHER INFORMATION: Escherichia coli), pHBA 1-hydroxylase (from Candida
; OTHER INFORMATION: parapsilosis), and UGT72B1 (from Arabidopsis thaliana).
US-10-462-162-47

Alignment Scores:
Pred. No.: 2.83e-101 Length: 3452
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-10-718-311-4 (1-165) x US-10-462-162-47 (1-3452)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 4 ATGTACACCCCGCTTAACGCAACTGCGTGGCTATTTAAAGAGATCCCTGCC 63
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 64 CTGGATCCGCAACTGCTGACGTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 123
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 124 CAGCAGGGGAAAAACGGTAAGCTGACGATGATCCCGAAGGTTGTGCGAGCAGATGAA 183
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 184 ATCCCCGAAGAACTGCCGCTGTCGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTTG 243
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 244 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGACCGTCTGTTCTGTCAACGTTA 303
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120

Db 952 TAAGCGGCGGAGCTGGCGTTACAAAAATTGGGTAAAAACGCCGTTAGGACGCTATCTGT 1011
QY 120 heThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 140
Db 1012 TCACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCCCGTGATGCCGGGCTGTGGG 1071
QY 140 lyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuP 160
Db 1072 GCGGACGTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTAC 1131
QY 160 roAlaSerProLeuTyr 165
Db 1132 CGGCGTCACCGTGTGTAC 1148

Search completed: May 30, 2006, 03:42:00
Job time : 1139.79 secs

THIS PAGE BLANK (USPTO)

Result No.	Query			Description		
	Score	Match	Length	DB	ID	
1	80	9.4	2907	7	US-11-293-697-1067	Sequence 1067, Ap
2	76.5	9.0	3550	7	US-11-293-697-509	Sequence 509, App
C	76	8.9	1104	7	US-11-217-529-2348	Sequence 2348, Ap
3	75.5	8.8	601	6	US-10-488-618-1930	Sequence 1930, Ap
4	75.5	8.8	1791	7	US-11-217-529-1938	Sequence 1938, Ap
5	74.5	8.7	2389	7	US-11-293-697-1066	Sequence 1066, Ap
6	73	8.5	4884	7	US-11-217-529-174332	Sequence 174332, A
C	73	8.5	1406	6	US-10-953-349-34874	Sequence 34874, A
8	72	8.4	3023	7	US-11-293-697-665	Sequence 665, App
9						

```
Qy 52 Arg-----GluGlyPhe-----ValGluGlnAenGluIleProGluGluLeuProLeu 67
Db 619 AGGGCTGTGCAGGGATTTCAAGGTACCTGTGCGAACCCCGAGCCCTCACTCATTTGAACCTCA 678
Qy 68 LeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCyeAlaAspGlyGluPro 87
Db 679 GTCAGAAACCAACCCCGTCAGTGGATTTAGAAA-----TGTGACAGGGAT--- 723
Qy 88 TrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeu 107
Db 724 ---TTCAGAGGCAATGGTCCAGCCC-----TCACAACCCAGCACCACTCTG 768
Qy 108 GlnLysLeuGlyLysThrProLeuGly-Arg-----TyrLeuPheThrSerThrLe 125
Db 769 CAGGCATGGAACACGAGGCCCCAGAGGACAGGGGCTTCTCAATATGGGCACACAGCACCA 828
Qy 125 uThrArgAspPheIleGluIle-----GlyArgAspAlaGlyLeuTr 139
Db 829 GACACTCCATAATAGAAAGTGAATAATCCCAAGCTCCCACTCCAGGAAGCTGACCACTG 888
Qy 139 p-----GlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGlu 156
Db 889 GGTGGATGATCCTCCCGCAGCGTCCAAACCTCTCTGGCAGTCCCTTGCTGCTCCAGGAG 946

RESULT 2
US-11-293-697-509/c
; Sequence 509, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; FILE REFERENCE: Novel full length cDNA
; CURRENT APPLICATION NUMBER: US/11/293,697
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 509
; LENGTH: 3550
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-509

Alignment Scores:
Pred. No.: 18.3 Length: 3550
Score: 76.50 Matches: 43
Percent Similarity: 41.1% Conservative: 15
Best Local Similarity: 30.5% Mismatches: 35
Query Match: 9.0% Indels: 49
DB: 7 Gaps: 8

US-10-718-311-4 (1-165) x US-11-293-697-509 (1-3550)
Qy 27 AspTrpLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 46
Db 397 GACAACACCCCTAAGTGAAGAGTGCCTCACAACAGAGTCCACAGAGCGGCAAAAT--GCG 340
Qy 47 SerValThrMetIleArgGluGlyPheValGluGln-----Asn 59
Db 339 AGTGTG-----GAGGAAGCATCTCCAGGATTCAGGGCGCCCGCTGACGAC 292
Qy 60 GluIleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIle 79
Db 291 TCCTGCCCGGCAAAAGTCCCGTCTCCAGCAGAATGAGACGATGCCCTGACCCATC 232
Qy 80 -----LeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValPro---- 96
Db 231 GTCAGGTTGACCTGC-----TGG-----CCAACCTGACTTCTCCAAGG 193
Qy 97 ValSerThrLeuSerGlyProGluLeuAla----- 106
```

```
Db 192 GTCACAATGCTCTTTGGTCTGAGTTACCCCTCATGGGAGACGAGAGGCATGGGAGGAAA 133
Qy 107 -----LeuGlnLysLeuGlyLysThrProLeu 115
Db 132 AGGGGGTACACACCCGAGGAGGAGTTGGCTTCACAGACAGAAAGACAAACCCGGGTG 73
Qy 116 GlyArgTyrLeuPhe-----ThrSerSerThrLeuThrArgAspPhe 129
Db 72 GGAAGATGGGGCTTTCTGCGGGGTAGGAATCACACACTTCGTCACACTGTGACAAAGACAGA 13
Qy 130 Ile 130
Db 12 ATT 10

RESULT 3
US-11-217-529-2348
; Sequence 2348, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 2348
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-2348

Alignment Scores:
Pred. No.: 3.89 Length: 1104
Score: 76.00 Matches: 42
Percent Similarity: 38.3% Conservative: 28
Best Local Similarity: 23.0% Mismatches: 63
Query Match: 8.9% Indels: 50
DB: 7 Gaps: 10

US-10-718-311-4 (1-165) x US-11-217-529-2348 (1-1104)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIle---Pro 19
Db 547 GTAGATAGACCCATAATTTCTTCAATCACCGGAGTAGCGTACTCTACAAGAGTAGTCCG 606
Qy 20 AlaLeuAspProGlnLeuLeuAspTrp-----LeuLeuLeuGlu 32
Db 607 TCACCTCCTAATAGATGATCGCGTCCGCAAAAAGATTGTAAGAGGACCTCAATGTGAG 666
Qy 33 AspSerMetThrLys-----ArgPheGluGlnGlnGlyLysThrVal-----Ser 47
Db 667 GTTAACAATAACAGCAGATGTTTGGAGAGGGCAAAATTCAGGTAAAGAGTCCAGGCTGGG 726
Qy 48 ValThrMetIleArgGluGly-----PheValGluGln----- 58
Db 727 ATAACCTTTGGTGGCTGAATCCACAGCAAAAAGTTGGAGCTATTTTCGCCGAAGATATTGG 786
Qy 59 -----AsnGluIleProGluLeuProLeuProLysGluSerArgTyrTrpLeu 76
Db 787 GATCAGCGCTCTATACCCGAGAACTTGGT-----GAAAAGTTGCTTACCGTTTATTA 840
Qy 77 ArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValPro 96
Db 841 GAAGAAATATCGAAAAGCGCT-----GCTGTCGGTAGAAGTCAACTTCCA 885
```

```
QY 97 ValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 116
Db 886 CTAGCA-----ATCGTATACATGCTATCGAAGAAAGATATCGGC 927
QY 117 ArgTyrLeuPheThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAla 136
Db 928 AGATTAAAGAAATTAAGACCGATAGATGAAGAGTTTATAACCTCTTGAGAGATATA 987
QY 137 GlyLeuTrpArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGlu 156
Db 988 -----AAGAGATCTTCTGCACAGAA 1008
QY 157 LeuPheLeu 159
Db 1009 GTCTTTTG 1017
RESULT 4
US-10-488-619-1930
; Sequence 1930, Application US/10488619
; Publication No. US20060099578A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P. C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; FILE REFERENCE: Physiological Conditions, And Genotyping Arrays
; CURRENT APPLICATION NUMBER: US/10/488,619
; CURRENT FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 3040
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1930
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-488-619-1930
Alignment Scores:
Pred. No.: 1.85 Length: 601
Score: 75.50 Matches: 32
Percent Similarity: 47.2% Conservative: 28
Best Local Similarity: 25.2% Mismatches: 38
Query Match: 8.8% Indels: 29
DB: 6 Gaps: 9
US-10-718-311-4 (1-165) x US-10-488-619-1930 (1-601)
QY 10 ArgAlaLeuArg-----TyrCysLysGluIleProAlaLeuAsp 22
Db 151 CGTTCCTTAAGAGAAAGCCATCAGAAAGTGTATTGCCACAGATCAAGGTCAACCACT 210
QY 23 ProGlnLeuLeuAsp-----TrpLeuLeuLeuGluAspSerMetThrLys 37
Db 211 GCTGAGCTTCTACCATACAGACCCATGGCTGGGCTCCAGCTTGTGATGATCTTTAAG 270
QY 38 ArgPheGluGlnGlnGlyThrValSer-----ValThrMetIleArgGluGly 54
Db 271 AATTTTAAAGGAAAGCAAAATCATCTCTGAATATGATCATCACCTTGATATGACT--- 327
QY 55 PheValGluGlnAsnGluIleProGluGluLeuProLeuProLysGluSerArgTyr 74
Db 328 TACATAGAGGAAATAAGCTACAGAAAGCTGTTCTGTAATTAAGAAAGGTA----- 378
QY 75 TrpLeuArgGluIle-----LeuLeuCysAlaAspGlyGluProTrpLeu 89
Db 379 ---CTGAGAGACATCGAGAGTCTCTCTGCACATAGCTGTGACAGGGGAAACAGGC--- 432
QY 90 AlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLys 109
Db 433 GCAGGGAAGTCCACTTTC---ATCAATACCTTGAGGGG-----GTGGGGCATGAAGAA 483
QY 110 LeuGlyLysThrProLeuGly 116
Db 484 AAAGGTGCAGCCCCCACTGGC 504
```

```
RESULT 5
US-11-217-529-1938
; Sequence 1938, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1938
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1938
Alignment Scores:
Pred. No.: 8.99 Length: 1791
Score: 75.50 Matches: 36
Percent Similarity: 34.3% Conservative: 10
Best Local Similarity: 26.9% Mismatches: 41
Query Match: 8.8% Indels: 47
DB: 7 Gaps: 8
US-10-718-311-4 (1-165) x US-11-217-529-1938 (1-1791)
QY 62 ProGluGluLeuPro-----LeuLeuPro 69
Db 94 CCAGATGACTTCCCGACGTGGTGAAGCTGTTAAGGAACACCGTTGGCAGCTGGAGCCC 153
QY 70 LysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeu 89
Db 154 GCATTGAGCAGGTATTTC-----GACGGGAA---TGGAGG 186
QY 90 AlaGly-----ArgThrValProValSerThrLeu 100
Db 187 GATGCTCCGAGGAGCGGATGGAGAGCCCACTGCACAGCCACGCCCAATCCACGCC 246
QY 101 SerGlyProGluLeuAlaLeuGlnLys-----LeuGlyLysThrProLeuGlyArgTyr 118
Db 247 ATGGCAGAGGAGTTGCGTGGACTCCGCCGCTGGTCCCGAGGCTCTATCA----- 300
QY 119 LeuPheThrSer-----SerThrLeuThrArgAspPhe 129
Db 301 ---TTTACAGCTGCATTCGCCGTAGTCAGACCACTTCCAGCAGAGCTTCGCTCATGATTTT 357
QY 130 IleGluIle-----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArg 146
Db 358 AGAACCATTTGGCTGAACGAGCGCTCCCAATACCGTATGTCATGTTCCAGTCTTCAGT 417
QY 147 LeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 418 TACAACGGTAACCCCTTTCTTTTGTGTTGCTGATCCCG 459
RESULT 6
US-11-293-697-1066
; Sequence 1066, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
```

```
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1066
; LENGTH: 2384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1066

Alignment Scores:
Pred. No.: 17.9 Length: 2384
Score: 74.50 Matches: 25
Percent Similarity: 48.3% Conservative: 4
Best Local Similarity: 41.7% Mismatches: 21
Query Match: 8.7% Indels: 10
DB: 7 Gaps: 2

US-10-718-311-4 (1-165) x US-11-293-697-1066 (1-2384)
Qy 65 LeuProLeu-LeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAs 84
Db 11 TTGCCACTCCCTTGGCAAGGAGCGGAGACTCAGTTCCCGGGGCTCTTTGGCGGCAC 70
Qy 84 pGlyGluProTrpLeuAlaGlyArgThrValValPro---ValSerThrLeuSerGlyPr 103
Db 71 AGGTGAGCCCTGGCTGCGCGCGCCCTCTCTCCCGCGGCTCCCAAGATGGGGGCTCC 130
Qy 103 oGluLeuAla-----LeuGlnLysLeuGlyLysThrPro 114
Db 131 GGAGTGGCGCCAGGCTCTGAGTACCTAGTCTGCAGACTAGCGGGCATTTGGCCA 188

RESULT 7
US-11-217-529-174332
; Sequence 174332, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 174332
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-174332

Alignment Scores:
Pred. No.: 2.76 Length: 489
Score: 73.00 Matches: 39
Percent Similarity: 34.3% Conservative: 7
Best Local Similarity: 29.1% Mismatches: 56
Query Match: 8.5% Indels: 32
DB: 7 Gaps: 6

US-10-718-311-4 (1-165) x US-11-217-529-174332 (1-489)
Qy 51 IleArgGluGlyPheValGluGlnAsnGluLeuProGluGluLeuPro----- 66
Db 31 TTAATAGATGCTCCAGGTATCATCAGCAATTTCCCGCCAGAGAGACTTGCACAGATGTGTGTGAGA 90
Qy 67 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGlu 78
```

```
Db 91 CTGTTGAGGAACCATGGTTGGCAGTTAGAACCAAGCATTTAGCCGATATTTTCGATGAGAG 150
Qy 79 IleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 98
Db 151 -----TGGAAAGCGCAACAGACAGATGGGAGAGTCC---ACTCAAACGTCT 195
Qy 99 ThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyr 118
Db 196 ACACCCATGGCTGAGACCTTGGTTCTCTCCCGCTTGGGACCTAGACCCGTTG----- 246
Qy 119 LeuPheThrSerSer-----ThrLeuThrArgAspPhe 129
Db 247 TTAATTACAGCTCAGTCCCGTAGTCAGACCATTCAGCAGAACTTCGCGAATGATTTC 306
Qy 130 IleGluLeu-----GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArg 146
Db 307 AGAACAAATGGTTAAATGGACGCTCTAAACACTGTATGTCCTATGTTGAATCTTTTAGT 366
Qy 147 LeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 367 TATGATGGTAACCCCTTCTCTTTTATTACTGTGATCCCT 408

RESULT 8
US-10-953-349-34874/C
; Sequence 34874, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34874
; LENGTH: 1406
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-34874

Alignment Scores:
Pred. No.: 12.7 Length: 1406
Score: 73.00 Matches: 29
Percent Similarity: 42.0% Conservative: 21
Best Local Similarity: 24.4% Mismatches: 35
Query Match: 8.5% Indels: 34
DB: 6 Gaps: 4

US-10-718-311-4 (1-165) x US-10-953-349-34874 (1-1406)
Qy 51 IleArgGluGlyPheValGluGlnAsnGluLeuPro-----GluGlu 64
Db 906 CTGAGGTCGAGCAGCGTGAGCTTCTCCAGCTTCGCGGCGGATGCAGGAGGAG 847
Qy 65 LeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuCysAlaAsp 84
Db 846 ATGCCCGCTGTTGAGGAGG----- 829
Qy 85 GlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGlu 104
Db 828 -----ATGAGCTCGTTGAGCGTGC CGGCC 805
Qy 105 LeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerSerThr 124
Db 804 ATGTCCCGATGTGTCGCGCAGCAGCGCTCACCCGCGGAAGGTTTC---GCCAGCAGC 748
Qy 125 LeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGlyArgSerArg 144
Db 747 ATCACC---GACCGCGTCGAGTTGTTGAAGTGGCGCGCACCGTGAACCTCGAAGTGGTGTG 691
Qy 145 LeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerPro 163
```

```
Db 690 TTCTTGATGAAAGCGCTGATCTTCTTGTGGAAGATGCCCGCCGACCTTGGCGG 634
RESULT 9
US-11-293-697-665
; Sequence 665, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 665
; LENGTH: 3023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-665
Alignment Scores:
Pred. No.: 50.6 Length: 3023
Score: 72.00 Matches: 35
Percent Similarity: 38.6% Conservatives: 16
Best Local Similarity: 26.5% Mismatches: 53
Query Match: 8.4% Indels: 28
DB: 7 Gaps: 5
US-10-718-311-4 (1-165) x US-11-293-697-665 (1-3023)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 448 TTAGAGACCCACATATAGTAGCTGACCACTGAGGTGATGTAATTAACAGGG 507
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 508 CTCATCTCTAAAAAACAACGCGCACTTGCTG-----CACCTTAGAGAACGATGGGAG 558
QY 41 GlnGln-----GlyLysThrValSerVal 48
Db 559 CAGCAGGTGTGCGCAGCAGATGGCAAACTGTGCGGCAAGAGGAAAGTGAACCCAG 618
QY 49 ThrMetIleArgGluGlyPheValGluGlnAsnGluLeuProGluGluLeuPro----- 66
Db 619 GCACCTCAGCTGAGCCATTCTCTAGGGGACTAATCATCTAGAGAGAACTGGCAGG 678
QY 67 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAsp 84
Db 679 AAAAGGCGAGAGCCCAAGGCAACAGAGCTGTGCGAAGAGTCTCTTAAACCCAGTGAC 738
QY 85 GlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGlu 104
Db 739 AATGAACAAGGCTTG-----CTGTGTTCTCCGGCTCTCCGCC----- 777
QY 105 LeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 116
Db 778 -----ATGAGAGTCTTTCATCCACAGTGAGGC 807
RESULT 10
US-10-524-433-1
; Sequence 1, Application US/10524433
; Publication No. US2006009676A1
; GENERAL INFORMATION:
; APPLICANT: Functional Genetics, Inc.
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN RAPAMYCIN
; FILE REFERENCE: 10784-023-228
; CURRENT APPLICATION NUMBER: US/10/524,433
; CURRENT FILING DATE: 2005-02-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
US-10-718-311-4 (1-165) x US-10-524-433-1 (1-966)
QY 39 PheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGln 58
Db 385 TTAAACACACAGGAAAG---TACACAGCCTCCCGCTCAGCGGAGCCAGTGGAAAGT 441
QY 59 AsnGluIleProGluGluLeuProLysGluSerArgTyrTrpLeuArgGlu 78
Db 442 AGCCACCTTCCTTGCATCTGCCACCTGGGACAGGAGGCTGTCACCTGCCGCGCTGG 501
QY 79 IleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 98
Db 502 GCTTTGG--TGC-----TGCTCTGTGACTCCACCTTGTCTGTGGTGCC 544
QY 99 ThrLeuSerGlyProGluLeuAlaGlnLysLeuGlyLysThrProLeuGlyArgTyr 118
Db 545 TCANGGAAGGT---CATCTCTNAAGAAACAAGCCACCTTCTCCCTCCCTGGCATG 601
QY 119 LeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeu 138
Db 602 CNTTTTTCNGCCACAGCCCGTGTGTGCCAACANAAATTTGGCNCNANGANGNAGGN 661
QY 139 TrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPhe 158
Db 662 TGGGANANAAACCCNAAACTA-----TTGTGG 688
QY 159 LeuProAlaSerPro 163
Db 689 CTNCCCTGTGTGCC 703
US-11-293-697-200/c
; Sequence 200, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 200
; LENGTH: 2490
; TYPE: DNA
```


THIS PAGE BLANK (uspto)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2006, 01:43:54 ; Search time 4259.96 Seconds
(without alignments)

3248.869 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MGHPPALQALRYCKEIPA.....RLSGKPLLLTEFLPASPLY 165

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/abs/ABSSWEB spool/US10718311/runat_26052006_164919_13483/app_query.fasta_1
-DB=EST_QPWT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02h
-USER=US10718311@CGN 1.1.7986 @runat_26052006_164919_13483 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -FGAPOP=10 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_est1: *
2: gb_est3: *
3: gb_est4: *
4: gb_est5: *
5: gb_est6: *
6: gb_hic: *
7: gb_est2: *
8: gb_est7: *
9: gb_est8: *
10: gb_est9: *
11: gb_gsa1: *
12: gb_gsa2: *
13: gb_gsa3: *
14: gb_gsa4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

C 1	618	72.4	832	13	CL666350 PRI0152b
2	508	59.5	828	13	CL687382 PRI0146c
3	262	30.7	536	14	AG192190 Pan trogl
4	236	27.6	533	14	AG192116 Pan trogl

ALIGNMENTS

RESULT 1
CL666350/c

LOCUS
DEFINITION

CL666350 PRI0152b_A03 - PRI0152b.B21 (832) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION
VERSION

CL666350 GI:50159025

KEYWORDS
SOURCE

GSS.
Pristionchus pacificus

ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

REFERENCE
AUTHORS

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satellite organism

TITLE

Pristionchus pacificus

JOURNAL
PUBMED

Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447

COMMENT

Contact: Sommer RJ
Evolutionary Biology

Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371
Fax: 00497071601498

Email: raif.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.

CC048367 01S0415-0
CL650764 PRI0110b
CL688660 PRI014a.D
AG192305 Pan trogl
BZ551713 pacis-60
BZ564576 pacis-164
AJ863670 Raletonia
BZ578851 msh2_6034
CL696699 PRI019d.B
BG799182 fp31c10.Y
BJ932163 BJ932163
DU782699 ASXB560.b
DY796666 III-952-7
CO384679 AGENCOURT
AV850332 Magnaport
BM870249 mgms009XO
CE278012 Eigr-g88-
CX288046 C01016B12
C0161709 FLD1_30.D
DA934706 DA934706
BQ642631 AGENCOURT
BQ650781 AGENCOURT
CK209417 FGAS02118
BJ578705 BJ578705
BJ566157 BJ566157
CL393085 ZMMBB019
CV120546 AGENCOURT
BU513771 AGENCOURT
CF513983 Cabud0007
BF543156 U1-R-AF1
BM265186 fw63f10.Y
BJ974078 BJ974078
BZ578446 msh2_5870
CV242171 WS02514.B
CL106163 ISB1-46F6
CG472449 ZMMBB028
AW236750 xms4e06.x
AW058694 fs0b08.Y
DA275761 DA275761
DB224823 DB224823
BC056456 Mus muscu

Seq primer: T7
Class: fosmid ends.

This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.

FEATURES
source
Location/Qualifiers

1..832
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

1..832
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Alignment Scores:
Pred. No.: 9,496-58 Length: 832
Score: 618.00 Matches: 120
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 72.4% Indels: 0
DB: 13 Gaps: 0

US-10-718-311-4 (1-165) x CL666350 (1-832)

Qy 46 ValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeu 65
Db 832 GTAACGGTACGATGATCGGAGGGTTTGTGAGCAGATGAATCCCGAAGACTG 773
Qy 66 ProLeuLeuProLyGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAaspGly 85
Db 772 CGCGTGTGCGAAGAGTCTCGTTACTGTTACGTGAAATTTTGTATGTCCGATGGT 713
Qy 86 GluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeu 105
Db 712 GAACCGTGGCTTGGCGGTGATACCGTCTGTCGTGCAACGTTAAGCGGCGGAGCTG 653
Qy 106 AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerSerThrLeu 125
Db 652 CGGTACAAATAATTGGTAAACGCCGTTAGACGCTATCTGTCACATCATCGACATTA 593
Qy 126 ThrArgAaspPheIleGluIleGlyArgAaspAlaGlyLeuTrpGlyArgSerArgLeu 145
Db 592 ACCGGGACTTATTGACATAGGCGGTGATGCCGGCTGTGGGGCGACGTTCCGCGCTG 533
Qy 146 ArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165
Db 532 CGATTACGCGTAACCGCTGTGCTAACAGAACTGTTTACCGCGGCGTCACCGTGTAC 473

RESULT 2
CL687382
LOCUS
DEFINITION
PR10146c_H03.2 - PR10146c.BR.(828) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CL687382.1 GI:50196135
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

1 (bases 1 to 828)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
Appab: an AcedB database for the nematode satellite organism
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de

FEATURES
source
Location/Qualifiers

1..828
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Alignment Scores:
Pred. No.: 1,53e-45 Length: 828
Score: 508.00 Matches: 98
Percent Similarity: 99.0% Conservative: 0
Best Local Similarity: 99.0% Mismatches: 1
Query Match: 59.5% Indels: 0
DB: 13 Gaps: 0

US-10-718-311-4 (1-165) x CL687382 (1-828)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 532 ATGTGCACACCCCGGTAAACGCAACTCGTCGCGCTATTTTAAAGACATCCCTGCC 591
Qy 21 LeuAaspProGlnLeuLeuAaspTrpLeuLeuGluAaspSerMetThrLysArgPheGlu 40
Db 592 CTGGATCGCAACTGCTCGACTGGCTGTGCTGGAGGATCCATGACAAAACGTTTGA 651
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAaspGlu 60
Db 652 CAGCAGGAAAAACCGTAAGCGTCACGATGATCCGGAAGGGTTTGTGACGACGAATGAA 711
Qy 61 IleProGluGluLeuProLeuLeuProLyGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 712 ATCCCCGAAGAACTGCGCGCTGCTCCGGAAGAGTCTCGTTACTGTTACGTGAAATTTTG 771
Qy 81 LeuCysAlaAaspGlyGluProTrpLeuAlaGlyArgThrValProValSerThr 99
Db 772 TTATGTCCGATGTTGAAACCGTGGCTGCGCGTGTACCGTCTCTCTGTGTCAACG 828

RESULT 3
AG192190
LOCUS

DEFINITION
Pan troglodytes DNA, clone: RP43-068J06.T7, genomic survey
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AG192190.1 GI:45224366
GSS.
Pan troglodytes (chimpanzee)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Pan.

REFERENCE
AUTHORS

1
Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.

BAC end sequences of Library RP-43
Unpublished
2 (bases 1 to 536)

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

1
Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
Direct Submission
Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
Biotechnology (KRIBB), Genome Research Center (GRC);
52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.krribb.re.kr, URL:http://phs.grc.krribb.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)
Clones are derived from the chimpanzee BAC library RP-43 This BAC

end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI

FEATURES

source

Location/Qualifiers
1..536
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-068J06.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"

ORIGIN

Alignment Scores:
Pred. No.: 1.67e-18 Length: 536
Score: 262.00 Matches: 51
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 30.7% Indels: 0
DB: 14 Gaps: 0
US-10-718-311-4 (1-165) x AG192190 (1-536)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|||||
Db 384 ATGTACACCCCGGTTAACGCAACTGCGTGGCGTCTATTGTAAGAGATCCCTGCC 443
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|||||
Db 444 CTGGATCCCACTGCTGCACCTGGCTGTGGAGGATTCATGACAAACGTTTGA 503
QY 41 GlnGlnGlyLysThrValSerValThrMetIle 51
|||||
Db 504 CAGCAGGAGAAACGGTAGCGTGACGATGATC 536

RESULT 4
AG192116
LOCUS
DEFINITION
AG192116 533 bp DNA linear GSS 06-MAR-2004
Pan troglodytes DNA, clone: RP43-068H07.T7, genomic survey
sequence.
ACCESSION
AG192116
VERSION
AG192116.1 GI:45224292
KEYWORDS
GSS.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.

REFERENCE
AUTHORS
Park, H., Kim, Y., Kim, S., Han, Y., Woo, T., Park, K., Eun, C.J.,
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.
BAC end sequences of Library RP-43
Unpublished
2 (bases 1 to 533)
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);
52, Oun-dong, Yuseong-gu, Daejeon 305-333, Korea
(E-mail: redstone@mail.krribb.re.kr, URL: http://phs.grc.krribb.re.kr/,
Tel: 82-42-866-7181, Fax: 82-42-860-4409)

COMMENTS
Clones are derived from the chimpanzee BAC library RP-43. This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.

PRIMERS
Sequencing: T7

LIBRARY

Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI

FEATURES

source

Location/Qualifiers
1..533
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-068H07.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"

ORIGIN

Alignment Scores:
Pred. No.: 1.27e-15 Length: 533
Score: 236.00 Matches: 47
Percent Similarity: 92.2% Conservative: 0
Best Local Similarity: 92.2% Mismatches: 4
Query Match: 27.6% Indels: 0
DB: 14 Gaps: 0
US-10-718-311-4 (1-165) x AG192116 (1-533)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|||||
Db 380 ATGTACACCCCGGTTAACGCAACTGCGTGGCGTCTATTGTAAGAGATCCCTGCC 439
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|||||
Db 440 CTGGATCCCACTGCTGCACCTGGCTGTGGAGGATTCATGACAAACGTTTGA 499
QY 41 GlnGlnGlyLysThrValSerValThrMetIle 51
|||||
Db 500 CAGCAGGAGAAACGGTTATCGTGACGATGATC 532

RESULT 5

CC048367

LOCUS

DEFINITION

CC048367

VERSION

CC048367.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

1 (bases 1 to 515)

Latehaw, S., Tan, B.-C., Settles, A.M. and McCarty, D.R.

Sequence tagged transposon insertions from the UniformMu maize
population

Unpublished (2003)

CONTACT: Donald R. McCarty

Plant Molecular and Cellular Biology Program

University of Florida

PO 110690 Gainesville, FL 32611-0690, USA

Tel: 352-392-1928 x322

Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:

01S-415-4

Class: transposon insertion site.

Location/Qualifiers

1..515

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="W22 (ACR, bz1-m9)"

/cultivar="UniformMu"

/db_xref="taxon:4577"

/clone="01S-415-4-7to12-E12"

/clone_lib="UniformMu MuTAIL Library"

FEATURES

source

/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

ORIGIN

Alignment Scores:
 Pred. No.: 1.2e-12 Length: 515
 Score: 209.00 Matches: 40
 Percent Similarity: 97.6% Conservative: 0
 Best Local Similarity: 97.6% Mismatches: 1
 Query Match: 24.5% Indels: 0
 DB: 12 Gaps: 0

US-10-718-311-4 (1-165) x CC048367 (1-515)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCyeLysGluLeuProAla 20
 Db 393 ATGTCACACCCCGCGTTAAGCAACTGCGTGGCGCTATTGTACAGAGATCCCTGCC 452
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
 Db 453 CTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 512
 Qy 41 Gln 41
 Db 513 CAA 515

RESULT 6
 CL650764 563 bp DNA linear GSS 09-JUL-2004
 LOCUS PR10110b.F06 - PR10110b.B21 (563) Mixed stage fosmid library of P.
 DEFINITION pacificus var. California Pristionchus pacificus genomic, genomic
 survey sequence.

ACCESSION CL650764.1 GI:50129323
 VERSION CL650764.1
 KEYWORDS GSS.

SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 563)
 Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
 AUTHORS AppaDB: an AcedB database for the nematode satellite organism
 TITLE Pristionchus pacificus

JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
 PUBMED 14681447

COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@uebingen.mpg.de
 This library was generated at Caltech, Pasadena, USA and end
 sequenced at Vancouver, Canada.
 Seq primer: T7
 Class: fosmid ends.
 Location/Qualifiers

FEATURES
 source

1..563
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus
 var. California"
 /note="Vector: pBpifos-5 Fosmid vector"

ORIGIN

Alignment Scores:

Pred. No.: 1.05e-09 Length: 563
 Score: 183.00 Matches: 38
 Percent Similarity: 97.4% Conservative: 0
 Best Local Similarity: 97.4% Mismatches: 0
 Query Match: 21.4% Indels: 1
 DB: 13 Gaps: 0

US-10-718-311-4 (1-165) x CL650764 (1-563)

Qy 128 AspPheileGluileGlyArgAspAlaGlyLeu-TrpGlyArgSerArgLeuArgLeu 147
 Db 2 GACITTTATTAGATAGCGCGTGATCGCGGCTCGTGGGGCGACGTTCCCGCTCGCAT 61
 Qy 147 uSerGlyAspProLeuLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165
 Db 62 AAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCGCGCTCACCGTTGTAC 116

RESULT 7

CL688660 838 bp DNA linear GSS 09-JUL-2004
 LOCUS PR1014a.D06.2 - PR1014a.BR (838) Mixed stage fosmid library of P.
 DEFINITION pacificus var. California Pristionchus pacificus genomic, genomic
 survey sequence.

ACCESSION CL688660.1 GI:50197738
 VERSION CL688660.1
 KEYWORDS GSS.

SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 838)
 Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
 AUTHORS AppaDB: an AcedB database for the nematode satellite organism
 TITLE Pristionchus pacificus
 JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
 PUBMED 14681447

COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@uebingen.mpg.de
 This library was generated at Caltech, Pasadena, USA and end
 sequenced at Vancouver, Canada.
 Seq primer: T7
 Class: fosmid ends.
 Location/Qualifiers

FEATURES
 source

1..838
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus
 var. California"
 /note="Vector: pBpifos-5 Fosmid vector"

ORIGIN

Alignment Scores:
 Pred. No.: 3.98e-08 Length: 838
 Score: 171.00 Matches: 36
 Percent Similarity: 94.7% Conservative: 0
 Best Local Similarity: 94.7% Mismatches: 1
 Query Match: 20.0% Indels: 1
 DB: 13 Gaps: 0

US-10-718-311-4 (1-165) x CL688660 (1-838)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyr-CysLysGluileProAla 20
 Db 725 ATGTCACACCCCGCGTTAAGCAACTGCGTGGCTGCTATTTTAAAGAGATCCCTGC 784
 Qy 20 aLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLys 37

```

Db      |||||||
785  CCTGGATCCGCAACTGCTCGACTGGCTGTGTGCTGGAGGATTCCATGACAAA 836
|||||

RESULT 8
AG192305
LOCUS   440 bp      DNA      linear      GSS 06-MAR-2004
DEFINITION Pan troglodytes DNA, clone: RP43-068M07.T7, genomic survey
sequence.
ACCESSION AG192305
VERSION   AG192305.1 GI:45224481
KEYWORDS  GSS.
SOURCE   Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Pan.
REFERENCE
AUTHORS  Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
          Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
          BAC end sequences of Library RP-43
          Unpublished
REFERENCE
AUTHORS  Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
          Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
          Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
          Bioscience and Biotechnology (KRIIB), Genome Research Center (GRC);
          52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
          (E-mail:redstone@mail.kribb.re.kr, URL:http://pns.grc.kribb.re.kr/,
          Tel:82-42-866-7181, Fax:82-42-860-4409)
          Clones are derived from the chimpanzee BAC library RP-43 This BAC
          end was generated during the R&D process and may have higher chance
          of clone tracking errors.
COMMENT  PRIMERS
          Sequencing: T7
          LIBRARY
          Vector : pBACe3.6
          R.Site 1 : EcoRI
          R.Site 2 : EcoRI.
          Location/Qualifiers
          1..440
             /organism="Pan troglodytes"
             /mol_type="genomic DNA"
             /db_xref="taxon:9598"
             /clone="RP43-068M07.T7"
             /sex="male"
             /cell_type="lymphocytes"
             /clone_lib="RP-43 Chimpanzee Male BAC Library"
ORIGIN
Alignment Scores:
Pred. No.: 2,64e-07 Length: 440
Score: 160.00 Matches: 30
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 18.7% Indels: 0
DB: 14 Gaps: 0
US-10-718-311-4 (1-165) x AG192305 (1-440)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysIysGluIleProAla 20
Db 349 ATGTACACCCCGCGTTAACGCAACTGCTGCGTATTGTTAAAGAGATCCCTGCC 408
QY 21 LeuAppProGlnLeuLeuAspTrpLeuLeu 30
Db 409 CTGGATCCGCAACTGCTCGACTGGCTGTTG 438
RESULT 9
BZ551713
LOCUS   1110 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pacs1-60_3287.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
pacs1-60_3287, genomic survey sequence.

```

```

ACCESSION BZ551713
VERSION   BZ551713.1 GI:27155321
KEYWORDS  GSS.
SOURCE   Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
          Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
          Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1110)
AUTHORS  Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
          Burns,J.L., Kaul,R. and Olsen,M.V.
          Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
          J. Bacteriol. (2002) In press
          Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.
          Location/Qualifiers
          1..1110
             /organism="Pseudomonas aeruginosa"
             /mol_type="genomic DNA"
             /strain="1-60"
             /db_xref="taxon:287"
             /clone="pacs1-60_3287"
             /note="clinical isolate 1-60 Whole genomic shotgun
             library."
ORIGIN
Alignment Scores:
Pred. No.: 1.28e-05 Length: 1110
Score: 150.00 Matches: 45
Percent Similarity: 37.7% Conservative: 12
Best Local Similarity: 29.8% Mismatches: 56
Query Match: 17.6% Indels: 38
DB: 11 Gaps: 3
US-10-718-311-4 (1-165) x BZ551713 (1-1110)
QY 21 LeuAppProGlnLeuLeuAspTrpLeuLeuLeuGluAAspSerMetThrLysArgPheGlu 40
Db 144 GTGGAACCCCTGCTGGAGGCTGGCAGACCCTGCGGACGAC-----GAA 188
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 189 TGCCAGGGGCTCGACGTC----- 206
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 207 -----CCTACCGGCAGTAGCGCTGGCTGCGCAGGCTCTAC 242
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 243 CTGCATGGCCACGACCGTCCCTGGGTGTCGCCCGCAGCGTGGCGGCGCGCCCTG 302
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 303 GAAGGCTCGGGCTTCACCTGGCGCTGCTGGCACCGCTCGCTGGCGAGTTGCTGTTTC 362
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIle----- 132
Db 363 AGCAGACGCGCTTCGAGCGCGGGCCCATCGAAGTCTGCCTATCCGCGCGCGCTCTG 422
QY 133 -----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGly 149
Db 423 CCCGCGGAGTCCCGCGGAGGTTCTTGGGGCCGCTGCTACGGTTTTCGCCGCGCG 482
QY 150 LysProLeuLeuLeuThrGluLeuPheLeuPro 160
          :::::

```

```

Db      483 CTCGGGTGCTGGTGGCGGAGTGACTACCG 515

RESULT 10
BZ564576/c
LOCUS   BZ564576       862 bp    DNA        linear    GSS 17-DEC-2002
DEFINITION   pacs2-164_4819.v2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION   pacs2-164_4819, genomic survey sequence.
VERSION     BZ564576.1  GI:27190626
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
ORGANISM    Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1 (bases 1 to 862)
AUTHORS     Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE       Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL     J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel.: 2062216954
            Fax: 2066857244
            Email: ckraymond@u.washington.edu
            Class: shotgun.
FEATURES   source
            Location/Qualifiers
                1..862
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /strain="2-164"
                /db_xref="taxon:287"
                /clone="pacs2-164_4819"
                /clone_lib="pacs2-164"
                /note="clinical isolate 2-164 Whole genomic shotgun
                library."
ORIGIN
Alignment Scores:
Pred. No.:      0.000191      Length:      862
Score:          138.00        Matches:     45
Percent Similarity: 37.3%      Conservative: 11
Best Local Similarity: 30.0%    Mismatches:  56
Query Match:    16.2%          Indels:       38
DB:             11             Gaps:         3

US-10-718-311-4 (1-165) x BZ564576 (1-862)

Qy      22 AspProGluLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGluGln 41
      ::::|||||
Db      662 AACCCCTCGCTGGAGGCGCTGGCAGACCCCTGGCGGAGC-----GAATGC 618
      :|||

Qy      42 GlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIle 61
      :|||||
Db      617 CAGGGGCTGACGTC----- 603

Qy      62 ProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeu 81
      :|||||
Db      602 -----CCTACCGGAGTAGCGGCTGGGTCGCGGAGGCTACCTG 564

Qy      82 CysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeuSer 101
      :|||
Db      563 CATGCCACGACCGTCCCTCGAGTGTTCGCCCGCAGCGTGGCGGCGCGAGCGCCCTGGAA 504

Qy      102 GlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThr 121
      :|||
Db      503 GGCTCGGGCTTCACCTGGCGGTGCTCGGACCCGCTCGTGGGCGAGTGTCTGTTGACG 444

Qy      122 SerSerThrLeuThrArgAspPheIleGluIle----- 132
      :|||
Db      443 GACAGCGCTTCGAGCGCGGGGCCCATCGAAGTCTGCCGCTATCCGGCGCGGTCGTGCC 384

```

```

Qy      133 -----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLys 150
      :|||
Db      383 GCCGAGGTCCCGCGCGAGGGTCTCTGGGGCGCTCGCTCACGGTTTCCCGCGCGCGCTC 324

Qy      151 ProLeuLeuLeuThrGluLeuPheLeuPro 160
      :|||
Db      323 GGGGTGCTGGTGGCGGAGGTGTACTACCG 294

RESULT 11
AJ863670/c
LOCUS   AJ863670       739 bp    DNA        linear    GSS 30-NOV-2005
DEFINITION   Ralstonia solanacearum GSS, clone V789R, genomic survey sequence.
ACCESSION   AJ863670
VERSION     AJ863670.1  GI:82937082
KEYWORDS    GSS; genome survey sequence.
SOURCE      Ralstonia solanacearum
ORGANISM    Ralstonia solanacearum
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Ralstonia.
REFERENCE   1 (bases 1 to 739)
AUTHORS     Munoz-Soriano,V., Arahall,D.R., Terol,J., Buades,C., Perez-Perez,A.,
            Llop,P., Belmonte,U.C.F., Lopez,M. and Perez-Alonso,M.
TITLE       Random genome sequencing of Ralstonia solanacearum strain IVIA 1602
            and comparative analysis with strain GM11000
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 739)
AUTHORS     Munoz-Soriano,V.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-2004) Genetica, Universidad de Valencia, C/ Dr.
            Moliner, 50, Burjassot, Valencia 46100, SPAIN
FEATURES   source
            Location/Qualifiers
                1..739
                /organism="Ralstonia solanacearum"
                /mol_type="genomic DNA"
                /strain="IVIA 1602"
                /db_xref="taxon:305"
                /clone="V789R"
ORIGIN
Alignment Scores:
Pred. No.:      0.0254      Length:      739
Score:          118.00        Matches:     42
Percent Similarity: 38.9%      Conservative: 23
Best Local Similarity: 25.1%    Mismatches:  64
Query Match:    13.8%          Indels:       38
DB:             14             Gaps:         3

US-10-718-311-4 (1-165) x AJ863670 (1-739)

Qy      25 LeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLys 44
      :|||||
Db      599 CTGTTGATGTTGGTGAAGATCTTGAATTCGCTGACGGCGGCTGCGTGGCGCTTCTCTCC 540
      :|||

Qy      45 ThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGlu 64
      :|||
Db      539 TCGTTCGCGTCAACTGCTCGGTGAG-----CAGCGGACC 504

Qy      65 LeuProLeu-----LeuProLysGluSerArgTyrTrpLeu 76
      :|||
Db      503 ATGCGCTGAGCAGCAGTAGTGGCTGCTGGGCTTGCCTGCGCGGAGACGTTGGCG 444

Qy      77 ArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValPro 96
      :|||
Db      443 CGCAAGTCTCTGCTGATCTGCGACGAAACCCCGGTGGTCTACGCACACACCATCGTGCAT 384

Qy      97 ValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 116
      :|||
Db      383 CCGCGCAGCTGGCGGCGGACTGGCCGCTTCTGAAGGCGCTGGGACCCAGCGCTCGGG 324

Qy      117 ArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGlu----- 131
      :|||
Db      323 CATCGCTGTTTCGCGGATCCGCGGCTTGCACGGGCGCTTCGAGTTTGTGCCAGCTGGAT 264

```

```

QY 132 -----1leGlyArg 134
Db 263 GTCCGGATCCGCTGTCAGCGCGGCCGTCGGCGCTGGCGCGCAGCGCGCTGGCGGC 204
QY 135 AspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeu 154
Db 203 ATGCGCGGCTGCGCGCAGCGCGCTTCGGTATTTCGCGCGGTCGCCAGCGCGATGCTGGTG 144
QY 155 ThrGluLeuPheLeuProLa 161
Db 143 ACGGAAGTGTTCGCTGCA 123

RESULT 12
LOCUS BZ578851
DEFINITION msh2_6034.x1 msh Pseudomonas aeruginosa genomic clone msh2_6034,
genomic survey sequence.
ACCESSION BZ578851
VERSION BZ578851.1 GI:27213912
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1141)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) in press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..1141
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_6034"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

ORIGIN
Alignment Scores:
Pred. No.: 1.31 Length: 1141
Score: 105.00 Matches: 44
Percent Similarity: 36.2% Conservative: 24
Best Local Similarity: 23.4% Mismatches: 70
Query Match: 12.3% Indels: 50
DB: 12 Gaps: 6

US-10-718-311-4 (1-165) x BZ578851 (1-1141)
QY 17 GluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThr 36
Db 466 CAACGTCCAGGCTCCCGCCACCGCTCTCGACTGCTTCGACGAAGGGTCGTGACC 525
QY 37 LysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVal 56
Db 526 CGCGCGC-----CTGACCGTCTCTCGCGCAGCGCGGTTC 558
QY 57 GluGlnAsnGluIleProGluGluLeuProLeuLeu----- 68
Db 559 CGCGTCGAACCCCTGTGTGAAGGCTGGCAGACCCCTGCGCAGCAGCAATGCCAGGGGCTC 618

```

```

QY 69 -----ProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGlu 86
Db 619 GAGCTCCCTCCCGCAGTAGCGCTGGGTCGCGAGGTCTACCTGTCATGCCACGACCGT 678
QY 87 ProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyPro----- 103
Db 679 CCTGGGTGNTGCCCGCAGCGTGGCGCGCGCAGCGCTGAAGGCTCGGCTCGACT 738
QY 104 -----GluLeuAlaLeuGlnLysLeuGlyLysThrPro 114
Db 739 TGGCGCTGCTCGCACCCGCTGTTGGCGAGTTGTTAGGAAAGCGCCTTGAGCGCGGC 798
QY 115 LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIle-----GluIle 132
Db 799 CCATCAAGCTGTGCGTATCCGCGCGCTTTCGCCGCCAGTTCGCCGGCAGGGTCTCTG 858
QY 133 GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArg-----Leu 147
Db 859 GGCGGTCCTTACGGTTTCCCGCGCGCTGGGTGCTGTGGCGAAGGTACCTACCGGGCTG 918
QY 148 SerGlyLysProLeuLeuLeuThr----- 155
Db 919 GGGACAGCCCGGAATTGCCGACGATAAACCCCGCGCTGAATTCGGAAGTCCCAAGTTTG 978
QY 156 GluLeuPheLeuProAlaSerPro 163
Db 979 AACCTTATCAAGCCTTTGTCCCT 1002

RESULT 13
LOCUS CL696699/c
DEFINITION CL696699 753 bp DNA linear GSS 10-JUL-2004
PRI019d_B03.2 - PRI019d_BR (753) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL696699
VERSION CL696699.1 GI:50218607
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 753)
AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
TITLE AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
PUBMED 14681447
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
FEATURES
Location/Qualifiers
1..753
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN
Alignment Scores:
Pred. No.: 9.34 Length: 753
Score: 95.00 Matches: 19
Percent Similarity: 100.0% Conservative: 0

```

```

Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      11.1%      Indels: 0
DB:              13         Gaps: 0

US-10-718-311-4 (1-165) x CL696699 (1-753)

QY 147 LeuSerGlyLysProLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165
DB 751 TTAAGCGTAACCGCTGTGCTAAAGAACTGTTTTTACCAGCGCTACCGTTGTAC 695

RESULT 14
LOCUS BG799182 591 bp mRNA linear EST 21-MAY-2001
DEFINITION fplacio.y1 zebrafish gridded kidney Danio rerio cDNA clone
IMAGE:4744722 5' similar to TR:Q9Y4G8 Q9Y4G8 KIAA0313 PROTEIN. ;,
mRNA sequence.
ACCESSION BG799182
VERSION BG799182.1 GI:14163514
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 581)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Persson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterson, R., and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
CONTACT Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA library Preparation: Leonard Ira Zon DNA Sequencing by:
Washington University Genome Sequencing Center Clone distribution:
Genome Systems, St. Louis, Missouri (web address:
www.genomesystems.com) (email contact: info@genomesystems.com) and
Research Genetics, Huntsville, Alabama (web address:
www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum Primardatenbank, Berlin, Germany (web address:
www.rzpd.de)
High quality sequence stop: 473.
FEATURES
source
1. 581
Location/Qualifiers
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:4744722"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XLOUR"
/clone_lib="zebrafish gridded kidney"
/note="Organ: kidney; Vector: pBK-CMV; Site 1: EcoRI;
Site 2: XhoI; Oligo dT cDNA library constructed from mRNA
pooled from pooled kidney tissue from 300 adult
zebrafish."

ORIGIN
Alignment Scores:
Pred. No.: 10.8 Length: 581
Score: 93.00 Matches: 38
Percent Similarity: 45.4% Conservative: 26
Best Local Similarity: 27.0% Mismatches: 43
Query Match: 10.9% Indels: 34
DB: 2 Gaps: 7

US-10-718-311-4 (1-165) x BG799182 (1-581)

QY 2 SerHisProAlaLeuThr-----GlnLeuArgAlaLeuArgTyrCysLysGluLeuPro 19
DB 117 AGTAACCGCTGACCTGGCACAGGCGCAGCAGCATCATCGACTACAGTACACGCCACCA 176

QY 20 AlaLeuAspProGlnLeuLeu-----AspTrpLeuLeu 30
DB 177 GAGTTGCAAGACCAGGTGCTCGGCATTTTCAAGCAGATCAGCAAGCCGTTACTTACTG 236

QY 31 LeuGluAspSerMetThrLysArg-----PheGluGlnGlnGlyLys 44
DB 237 GCCAATGAGACACACACAGCCAGAGAGCTGCCAATCTAGCCATAAAGAGAGTTTGGCTTG 296

QY 45 ThrValSer-----ValThrMetIleArgGluGlyPheVal 56
DB 297 TCTCGAGTCCCGAAGCTTTTTTCGCTCTGTGGAAGTTCAGTCACACAGGAAGGAGTTATC 356

QY 57 GluGlnAenGluLeuProGluLeuProLeuProLys-----GluSer 72
DB 357 AAGCAGACAGCACTGCCTGATCAACTGTCCAACTGGCTGACAGATCCAACTGAGTGCC 416

QY 73 ArgTyrTrpLeuArgGluLeuLeu-----LeuCysAlaAspGlyGluProTrp 88
DB 417 AGGTACTATCTAAAGAGCAATATGAAACCGAGACGCTGTGTCGGATGTGGAAGCACTG 476

QY 89 ---LeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeu 107
DB 477 GAGATGAGACAGAGAATCTGTGTCGCCCTGTTGCTTTGAGCTCAATGGAGATAGCCCAAT 536

QY 108 Gln 108
DB 537 CAG 539

RESULT 15
LOCUS BG7932163 666 bp mRNA linear EST 07-OCT-2005
DEFINITION BG7932163 Daphnia magna cDNA library Daphnia magna cDNA clone
IMAGE:0004_L04.r.3', mRNA sequence.
ACCESSION BG7932163
VERSION BG7932163.1 GI:66942524
KEYWORDS EST.
SOURCE Daphnia magna
ORGANISM Daphnia magna
Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda;
Diplostroica; Cladocera; Anomopoda; Daphniidae; Daphnia.
Watanabe, H., Tatarazako, N., Oda, S., Nishide, H., Uchiyama, I.,
Morita, M. and Iguchi, T.
Analysis of expressed sequence tags of the water flea Daphnia magna
Genome 48 (4), 606-609 (2005)
16094427
Contact: Hajime Watanabe
Okazaki Institute for Bioscience
National Institute of Natural Sciences
Higashiyama 5-1, Myodaiji, Okazaki, Aichi, 444-8787, Japan
Tel: 81-564-59-5237
Fax: 81-564-59-5236
Email: watanabe@nibb.ac.jp.
FEATURES
source
1. 666
Location/Qualifiers
/organism="Daphnia magna"
/mol_type="mRNA"
/db_xref="taxon:35525"
/clone="WITH001_0004_L04.r"
/sex="female"
/tissue_type="whole body"
/dev stage="adult"
/clone_lib="Daphnia magna cDNA library"

ORIGIN
Alignment Scores:
Pred. No.: 21.8 Length: 666
Score: 91.00 Matches: 29

```

Percent Similarity: 37.5% Conservative: 13
Best Local Similarity: 25.9% Mismatches: 24
Query Match: 10.7% Indels: 46
DB: 2 Gaps: 5

US-10-718-311-4 (1-165) x BJ932163 (1-666)

```
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 312 GTTCCAGACCATATAGCTTACTT-----TATGGCTTCGTGCTGCGCTT 356
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 357 TCTTCTTCAACGCGCAGGCGCTGG----- 380
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 -----AATAAGTGCCCTCATTCTCGTTATTACTTT 410
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 ThrSerSerThrLeuThrArg-----Asp 128
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 411 TAAGGATTGCACATATCTCGTAGAACGCAACCAACCTGGATGCAAGTTATTTTTC 470
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 PheIleGluIleGlyArgAspAla-----GlyLeuTrpGlyArgArg 142
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 471 TTCCTTGATTGGCGCGTACGGCGCATGGCGAGTTCGCTTACGGGCGTCATGTCATTATA 530
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 143 SerArgLeuArgLeu---SerGlyLysProLeuLeu 153
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 531 AGTCGATTAGTTGTGCCAGTTTCACGTCCAATGCTG 566
Db   ::::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 30, 2006, 04:08:20
Job time : 4262.96 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2006, 01:01:59 ; Search time 4098 Seconds
(without alignments)
10673.508 Million cell updates/sec

Title: US-10-718-311-7
Perfect score: 684
Sequence: 1 atgggttcctctgtcatttc.....cggcgtcaccgttgactaa 684

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_env.*
2: gb_pat.*
3: gb_ph.*
4: gb_pl.*
5: gb_pr.*
6: gb_ro.*
7: gb_srs.*
8: gb_sy.*
9: gb_un.*
10: gb_vi.*
11: gb_ov.*
12: gb_hcg.*
13: gb_in.*
14: gb_om.*
15: gb_ba.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	684	100.0	684	2	AR428736 Sequence
2	684	100.0	684	2	AR455412 Sequence
3	684	100.0	684	2	AX329368 Sequence
4	684	100.0	684	2	AX382258 Sequence
5	498	72.8	854	2	AR144788 Sequence
6	498	72.8	1498	15	M96268 Escherichia
7	498	72.8	2000	2	E11274 ubiC and ub
8	498	72.8	2000	15	M93136 Escherichia
9	498	72.8	2000	15	M93413 Escherichia
10	498	72.8	2034	15	X57434 E.coli ubiC
11	498	72.8	2348	15	X66619 E.coli gene
12	498	72.8	6641	2	AR144789 Sequence
13	498	72.8	110000	15	Continuation (43 o
14	498	72.8	110000	15	Continuation (43 o
15	498	72.8	176195	15	U00005 E.coli chr
16	496.4	72.6	1396	15	DQ087228 Escherich
17	495	72.4	495	2	A47932 Sequence 1
18	495	72.4	495	2	AR428733 Sequence

19	495	72.4	495	2	AR455409 Sequence
20	495	72.4	495	2	AX329364 Sequence
21	495	72.4	495	2	AX382254 Sequence
C	494.8	72.3	110000	15	Continuation (42 o
23	490	71.6	110000	15	AE005174_51
24	488.4	71.4	110000	15	BA000007_50
25	488.4	71.4	110000	15	BA000007_51
C	488.4	71.4	110000	15	Continuation (52 o
26	488.4	71.4	110000	15	CP000034_42
27	488.4	71.4	110000	15	CP000038_44
28	485.2	70.9	110000	15	AE014075_47
29	481.2	70.4	1318	2	E07849
C	478.8	70.0	110000	15	Continuation (48 o
30	478.8	70.0	110000	15	AE005674_43
31	478.8	70.0	110000	15	Continuation (44 o
32	320.8	46.9	23880	15	AE014073_34
33	320.8	46.9	23880	15	AE008898
34	319.2	46.7	110000	15	AE014613_42
35	319.2	46.7	110000	15	AE017220_43
36	319.2	46.7	110000	15	CP000026_41
37	319.2	46.7	110000	15	CP000026_42
38	282	41.2	645	2	AL627282
39	184	26.9	599	4	AR384505 Sequence
40	184	26.9	742	4	M13543 Tomato (L.e
41	184	26.9	1097	4	M15236 Tomato RuBP
42	179	26.2	2776	4	BT013023 Lycopersi
43	146.8	21.5	3323	4	X05983 Tomato rbcS
44	145.6	21.3	806	4	X69759 S.tuberosum
45	143.6	21.0	1054	4	AY220079 Nicotiana
					X05985 Tomato rbcS

ALIGNMENTS

RESULT 1	AR428736	Sequence 16 from patent US 6642036.	684 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR428736					
DEFINITION	Sequence 16 from patent US 6642036.					
ACCESSION	AR428736					
VERSION	AR428736.1	GI:40188466				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 684)					
AUTHORS	Flint, D., Meyer, K. and Viitanen, P.					
TITLE	Sinapoylglucose:malate sinapoyltransferase form malate conjugates from benzoic acid glucosides					
JOURNAL	Patent: US 6642036-A 16 04-NOV-2003;					
	E. I. Du Pont de Nemours and Company; Wilmington;					
	WOX;					
FEATURES	Location/Qualifiers					
source	1..684					
	/organism="unknown"					
	/mol_type="genomic DNA"					

ORIGIN	Query Match	100.0%;	Score 684;	DB 2;	Length 684;
	Best Local Similarity	100.0%;	Pred. No. 6.8e-202;		
	Matches 684;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGGCTTCTCTGTGTCATTTCTTTCAGCAGCTGTTGGCACACGCGAGCAATGTTACACAAGCT	60		
Db	1	ATGGCTTCTCTGTGTCATTTCTTTCAGCAGCTGTTGGCACACGCGAGCAATGTTACACAAGCT	60		
QY	61	AGCATGGTGGACCTTTCATCTGCTCAATCTTCAGCCACTTTCCTCTTACAAAGAAG	120		
Db	61	AGCATGGTGGACCTTTCATCTGCTCAATCTTCAGCCACTTTCCTCTTACAAAGAAG	120		
QY	121	CAAAACCTTGACATCACCCTTCCATTCGTAGCAATGGTGAAGAGTTAGCTGCATGCAGGTG	180		
Db	121	CAAAACCTTGACATCACCCTTCCATTCGTAGCAATGGTGAAGAGTTAGCTGCATGCAGGTG	180		
QY	181	TGGCATATGTCTACACCCCGGTTAAACGAACTCGCGTGGCTCGCTGCTATTGTAAAGAGATC	240		

Db 181 TGGCATATGTCACACCCCGGTTAAGCACTGCGTGGCTGCGCTATTGTTAAAGAGATC 240
Qy 241 CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300
Db 241 CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300
Qy 301 TTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTCTGAGCAG 360
Db 301 TTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTCTGAGCAG 360
Qy 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Db 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Qy 421 ATTTTGTATTGTGCCGATGGTGAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 480
Db 421 ATTTTGTATTGTGCCGATGGTGAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 480
Qy 481 ACGTTAAGCGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACGCCGTTAGGACGCTAT 540
Db 481 ACGTTAAGCGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACGCCGTTAGGACGCTAT 540
Qy 541 CTGTTACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 600
Db 541 CTGTTACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 600
Qy 601 TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Db 601 TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Qy 661 TTACCGGCGTCACCGTTGTACTAA 684
Db 661 TTACCGGCGTCACCGTTGTACTAA 684

RESULT 2
AR455412 AR455412 684 bp DNA linear PAT 20-FEB-2004
LOCUS Sequence 7 from patent US 6683231.
DEFINITION AR455412
ACCESSION AR455412
VERSION AR455412.1 GI:42689960
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 684)
AUTHORS Meyer, K., Viitanen, P.V. and Van Dyk, D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 7 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES
source
1..684
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 684; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 6.8e-202;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGCTTCTCTGTGCTATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCTCTGTGCTATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Qy 61 AGCATGGTTGACACCTTTCACCTGGTCTCAAACTCTTCAGCCACTTTCCCTGTTTACAAAGAG 120
Db 61 AGCATGGTTGACACCTTTCACCTGGTCTCAAACTCTTCAGCCACTTTCCCTGTTTACAAAGAG 120
Qy 121 CAAAACCTTGACATCATCTTCATGCTAGCAATGGTGGAGAGTTAGCTGATGAGGTG 180
Db 121 CAAAACCTTGACATCATCTTCATGCTAGCAATGGTGGAGAGTTAGCTGATGAGGTG 180
Qy 181 TGGCATATGTCACACCCCGGTTAAGCACTGCGTGGCTGCGCTATTGTTAAAGAGATC 240

Db 181 TGGCATATGTCACACCCCGGTTAAGCACTGCGTGGCTGCGCTATTGTTAAAGAGATC 240
Qy 241 CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300
Db 241 CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300
Qy 301 TTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTCTGAGCAG 360
Db 301 TTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTCTGAGCAG 360
Qy 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Db 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Qy 421 ATTTTGTATTGTGCCGATGGTGAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 480
Db 421 ATTTTGTATTGTGCCGATGGTGAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 480
Qy 481 ACGTTAAGCGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACGCCGTTAGGACGCTAT 540
Db 481 ACGTTAAGCGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACGCCGTTAGGACGCTAT 540
Qy 541 CTGTTACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 600
Db 541 CTGTTACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGGTGCTGCTGTTCTCTGTGCA 600
Qy 601 TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Db 601 TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Qy 661 TTACCGGCGTCACCGTTGTACTAA 684
Db 661 TTACCGGCGTCACCGTTGTACTAA 684

RESULT 3
AX329368 AX329368 684 bp DNA linear PAT 08-JAN-2002
LOCUS Sequence 7 from Patent WO0194607.
DEFINITION AX329368
ACCESSION AX329368
VERSION AX329368.1 GI:18102367
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Meyer, K., van Dyk, D.E. and Viitanen, P.V.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: WO 0194607-A 7 13-DEC-2001;
E.I. Dupont De Nemours (US)
FEATURES
source
1..684
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="synthetic CPL"

ORIGIN
Query Match 100.0%; Score 684; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 6.8e-202;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGCTTCTCTGTGCTATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCTCTGTGCTATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Qy 61 AGCATGGTTGACACCTTTCACCTGGTCTCAAACTCTTCAGCCACTTTCCCTGTTTACAAAGAG 120
Db 61 AGCATGGTTGACACCTTTCACCTGGTCTCAAACTCTTCAGCCACTTTCCCTGTTTACAAAGAG 120
Qy 121 CAAAACCTTGACATCATCTTCATGCTAGCAATGGTGGAGAGTTAGCTGATGAGGTG 180

Db 121 CAAACCTTGACATCACTCCATTGCTAGCAATGGTGAAGAGTTAGCTGCATGCAAGTG 180
Qy 181 TGGCATATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATC 240
Db 181 TGGCATATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATC 240
Qy 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGATTCATGACAAACGT 300
Db 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGATTCATGACAAACGT 300
Qy 301 TTTGACACAGGGAAGAAACGTTAGCTGACGATGATCCGGAAGGTTTGTGAGCAG 360
Db 301 TTTGACACAGGGAAGAAACGTTAGCTGACGATGATCCGGAAGGTTTGTGAGCAG 360
Qy 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGGTTACGTGAA 420
Db 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGGTTACGTGAA 420
Qy 421 ATTTTGTATGTGCGGATGGTGAAACCGTGGCTTGGCGGTGTAACCGTCTCTGTGTCA 480
Db 421 ATTTTGTATGTGCGGATGGTGAAACCGTGGCTTGGCGGTGTAACCGTCTCTGTGTCA 480
Qy 481 AGTTAAGCGGCGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTAT 540
Db 481 AGTTAAGCGGCGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTAT 540
Qy 541 CTGTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGGTGATGCCGGCTG 600
Db 541 CTGTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGGTGATGCCGGCTG 600
Qy 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660
Db 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660
Qy 661 TTACCGGCTCACCGTTGACTAA 684
Db 661 TTACCGGCTCACCGTTGACTAA 684

RESULT 4
AX382258 AX382258 684 bp DNA linear PAT 18-MAR-2002
LOCUS Sequence 16 from Patent WO0204653.
DEFINITION
ACCESSION AX382258
VERSION AX382258.1 GI:19577037
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

1
REFERENCE Flint, D., Meyer, K. and Viitanen, P.V.
AUTHORS Sinapolyglucose:maltate sinapolytransferase form maltate conjugates
TITLE from benzoic acid glucosides
JOURNAL Patent: WO 0204653-A 16 17-JAN-2002;
E.I. DUPONT DE NEUMOURS AND COMPANY (US)
FEATURES Location/Qualifiers
source
1. 684
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="open reading frame of the chloroplast-targeted CPL fusion protein"

ORIGIN
Query Match 100.0%; Score 684; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 6.8e-202;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGGTTCTCTGTCATTCTTTAGCAGCTGTTGCCACACGAGCAATGTTACACAGCT 60
Db 1 ATGGGTTCTCTGTCATTCTTTAGCAGCTGTTGCCACACGAGCAATGTTACACAGCT 60
Qy 61 AGCATGTTGCACTTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTTACAAAGAAG 120

Db 61 AGCATGTTGCACTTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTTACAAAGAAG 120
Qy 121 CAAACCTTGACATCACTCCATTGCTAGCAATGGTGAAGAGTTAGCTGCATGCAAGTG 180
Db 121 CAAACCTTGACATCACTCCATTGCTAGCAATGGTGAAGAGTTAGCTGCATGCAAGTG 180
Qy 181 TGGCATATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATC 240
Db 181 TGGCATATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATC 240
Qy 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGATTCATGACAAACGT 300
Db 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGATTCATGACAAACGT 300
Qy 301 TTTGACACAGGGAAGAAACGTTAGCTGACGATGATCCGGAAGGTTTGTGAGCAG 360
Db 301 TTTGACACAGGGAAGAAACGTTAGCTGACGATGATCCGGAAGGTTTGTGAGCAG 360
Qy 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGGTTACGTGAA 420
Db 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGGTTACGTGAA 420
Qy 421 ATTTTGTATGTGCGGATGGTGAAACCGTGGCTTGGCGGTGTAACCGTCTCTGTGTCA 480
Db 421 ATTTTGTATGTGCGGATGGTGAAACCGTGGCTTGGCGGTGTAACCGTCTCTGTGTCA 480
Qy 481 AGTTAAGCGGCGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTAT 540
Db 481 AGTTAAGCGGCGCGGAGCTGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTAT 540
Qy 541 CTGTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGGTGATGCCGGCTG 600
Db 541 CTGTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGGTGATGCCGGCTG 600
Qy 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660
Db 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660
Qy 661 TTACCGGCTCACCGTTGACTAA 684
Db 661 TTACCGGCTCACCGTTGACTAA 684

RESULT 5
AX382258 AX382258 854 bp DNA linear PAT 08-AUG-2001
LOCUS Sequence 24 from patent US 6210937.
DEFINITION
ACCESSION AX382258
VERSION AX382258.1 GI:15106655
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 854)
AUTHORS Ward, T.E., Watkins, C.S., Bulmer, D.K., Johnson, B.F. and Amarutunga, M.
TITLE Development of genetically engineered bacteria for production of selected aromatic compounds
JOURNAL Patent: US 6210937-A 24 03-APR-2001;
FEATURES Location/Qualifiers
source
1. 854
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 72.8%; Score 498; DB 2; Length 854;
Best Local Similarity 100.0%; Pred. No. 9.5e-144;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 187 ATGTCACACCCCGGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATTCCTGCC 246
Db 321 ATGTCACACCCCGGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATTCCTGCC 380

```
QY 247 CTGGATCCGCAACTGCTGACATGGCTGTTGCTGGAGGATTCATCACAAAACGTTTGA 306
Db |||||||
QY 381 CTGGATCCGCAACTGCTGACATGGCTGTTGCTGGAGGATTCATCACAAAACGTTTGA 440
Db |||||||
QY 307 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCCGAAGGTTTGTGAGCAGAAATGAA 366
Db |||||||
QY 441 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCCGAAGGTTTGTGAGCAGAAATGAA 500
Db |||||||
QY 367 ATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426
Db |||||||
QY 501 ATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 560
Db |||||||
QY 427 TTATGTCCGATGGTAACCGTGGCTGCCGTCGTACCGTCTGTCCTGTGTCAACGTTA 486
Db |||||||
QY 561 TTATGTCCGATGGTAACCGTGGCTGCCGTCGTACCGTCTGTCCTGTGTCAACGTTA 620
Db |||||||
QY 487 AGCGGGCCGGAGCTGGCTTACAAAATTTGGGTAATAACGCCGTTAGGACGCTATCTGTTT 546
Db |||||||
QY 621 AGCGGGCCGGAGCTGGCTTACAAAATTTGGGTAATAACGCCGTTAGGACGCTATCTGTTT 680
Db |||||||
QY 547 ACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGGTGATGCCGGGCTGTGGGG 606
Db |||||||
QY 681 ACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGGTGATGCCGGGCTGTGGGG 740
Db |||||||
QY 607 CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 666
Db |||||||
QY 741 CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 800
Db |||||||
QY 667 CGGTCACCGTTGTACTAA 684
Db |||||||
QY 801 CGGTCACCGTTGTACTAA 818
Db |||||||
```

```
RESULT 6
LOCUS ECOUBIA 1498 bp DNA linear BCT 28-MAR-1994
DEFINITION Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA)
ACCESSION M96268
VERSION 1
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; aerobic respiratory
deficiency; chorismate lyase; membrane-bound protein;
prenyltransferase; transferase; ubiA gene; ubiC gene; ubiquinone
biosynthesis.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1498)
AUTHORS Wu,G., Williams,H.D., Gibson,F. and Poole,R.K.
TITLE Mutants of Escherichia coli affected in respiration: the cloning
and nucleotide sequence of ubiA, encoding the membrane-bound
p-hydroxybenzoate:octaprenyltransferase
J. Gen. Microbiol. 139 (Pt 8), 1795-1805 (1993)
JOURNAL 8409922
PUBMED Location/Qualifiers
FEATURES
source
1..1498
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/strain="K-12"
/db_xref="taxon:562"
/map="91.8 min"
/tissue_lib="cosmid"
<1..524
/feature="ORF; the sequence before nucleotide number 110
might not be genuine. Three Clai sites, which is the
enzyme used for cloning, were identified in this region,
so it could be a cloning artifact"
/codon_start=3
/transl_table=11
/protein_id="AAAL7026.1"
/db_xref="GI:347887"
```

CDS

```
/translation="RFSEVNFDTQOMIDFINRENTKWLIDKNNFFIQLIDQALRST
DDMIKANVWHLKYEIRSDDDVSPFIETEDNLRTPNTNELLTENDNIFILFSSVDGVP
MVSSQRJLHMLNPTKDTNNMNSTYYIKSRHEMLPVNLTQETLFSKSHGKALFPFIT
ASWAHRIMNKGV"
735..738
RBS
gene
747..1244
/gene="ubiC"
747..1244
CDS
/gene="ubiC"
/codon_start=1
/transl_table=11
/product="chorismate lyase"
/protein_id="AAAL7026.1"
/db_xref="GI:347888"
/translation="MSPHPALTLQALRYCKEIPALDPOLLWLLELLEDSMTKRFBOQKG
TVSTMIREGFEQNEIPEELPLPKESRYLWREILLCADGEPWLAGRTVVPVSTLSG
PELAQLQKGLTPLGRLYFTSLTRDFIEIGRDAGLWGRSRLSLSGKPLLLTELFLP
ASPLY"
1257..1498
gene
/gene="ubiA"
1257..1498
CDS
/gene="ubiA"
/note="The last part of this ORF is to be found in entry
K00127.; ORF"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate-octaprenyl transferase"
/protein_id="AAAL7026.1"
/db_xref="GI:347889"
/translation="MEWSLTQNKLLAFHRLMRTDKPIGALLLLWPTLWALWVATRGVP
QLWLAVFVAGVWLMRAAGCVVNDYADRKPDGHVKT"
ORIGIN
```

```
Query Match 72.8%; Score 498; DB 15; Length 1498;
Best Local Similarity 100.0%; Pred. No. 1.1e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTCACACCCCGGTTAAACGCACTGCGTGGCTATTGTAAAGAGATCCCTGCC 246
Db |||||||
QY 747 ATGTCACACCCCGGTTAAACGCACTGCGTGGCTATTGTAAAGAGATCCCTGCC 806
Db |||||||
QY 247 CTGGATCCGCAACTGCTGACATGGCTGTTGCTGGAGGATTCATCACAAAACGTTTGA 306
Db |||||||
QY 807 CTGGATCCGCAACTGCTGACATGGCTGTTGCTGGAGGATTCATCACAAAACGTTTGA 866
Db |||||||
QY 307 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCCGAAGGTTTGTGAGCAGAAATGAA 366
Db |||||||
QY 867 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCCGAAGGTTTGTGAGCAGAAATGAA 926
Db |||||||
QY 367 ATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426
Db |||||||
QY 927 ATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 986
Db |||||||
QY 427 TTATGTCCGATGGTAACCGTGGCTGCCGTCGTACCGTCTGTCCTGTGTCAACGTTA 486
Db |||||||
QY 987 TTATGTCCGATGGTAACCGTGGCTGCCGTCGTACCGTCTGTCCTGTGTCAACGTTA 1046
Db |||||||
QY 487 AGCGGGCCGGAGCTGGCGTTACAAAATTTGGGTAATAACGCCGTTAGGACGCTATCTGTTT 546
Db |||||||
QY 1047 AGCGGGCCGGAGCTGGCGTTACAAAATTTGGGTAATAACGCCGTTAGGACGCTATCTGTTT 1106
Db |||||||
QY 547 ACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGGTGATGCCGGGCTGTGGGG 606
Db |||||||
QY 1107 ACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGGTGATGCCGGGCTGTGGGG 1166
Db |||||||
QY 607 CGAGGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 666
Db |||||||
QY 1167 CGAGGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 1226
Db |||||||
QY 667 CGGTCACCGTTGTACTAA 684
Db |||||||
QY 1227 CGGTCACCGTTGTACTAA 1244
Db |||||||
```

```
RESULT 7
LOCUS      E11274
DEFINITION ubiC and ubiA gene.
ACCESSION  E11274
VERSION    E11274.1 GI:22024916
KEYWORDS   JP 1996107789-A/1.
SOURCE     Escherichia coli
ORGANISM   Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE  1 (bases 1 to 2000)
AUTHORS    Matsuda.H., Kawamuki.M. and Nakagawa.T.
TITLE      PRODUCTION OF UBIQUINONE-10
JOURNAL    Patent: JP 1996107789-A 1 30-APR-1996;
ALPHA- SHOKUHIN KK
COMMENT    OS Escherichia coli
          FN JP 1996107789-A/1
          PD 30-APR-1996
          PF 13-OCT-1994 JP 1994273071
          PI MATSUDA HIDEYUKI, KAWAMUKI MAKOTO, NAKAGAWA TSUYOSHI PC
          C12N15/09,A61K31/12,A61K31/12,C07H21/04,C12N1/21,C12P7/66, PC
          (C12N1/21,
          PC C12R1:01), (C12P7/66,C12R1:01);
          CC strandedness: Double;
          CC topology: Linear;
          FH Key Location/Qualifiers
          FH source 1..2000
          FT /organism='Escherichia coli'
          FT CDS 380..877
          FT /product='UbiC protein'
          FT CDS 890..1762
          FT /product='UbiA protein'.
          FT Location/Qualifiers
          FT source 1..2000
          FT /organism='Escherichia coli'
          FT /mol_type='genomic DNA'
          FT /db_xref='taxon:562'

FEATURES
source
1..2000
/organism='Escherichia coli'
/mol_type='genomic DNA'
/db_xref='taxon:562'

ORIGIN
Query Match 72.8%; Score 498; DB 2; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTACACCCGCGTTAAGCACTGCGTGGCTCGCTATGTAAGAGATCCCTGCC 246
DB 187 |||||
DB 380 ATGTACACCCGCGTTAAGCACTGCGTGGCTCGCTATGTAAGAGATCCCTGCC 439
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 306
DB 247 |||||
DB 440 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 499
QY 307 CAGCAGGGAAAAACGGTAGCGTAGCATGATCCGGAAGGGTTTGTGAGCAGAGATGAA 366
DB 307 |||||
DB 500 CAGCAGGGAAAAACGGTAGCGTAGCATGATCCGGAAGGGTTTGTGAGCAGAGATGAA 559
QY 367 ATCCCGGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGGTAGTGAAATTTTG 426
DB 367 |||||
DB 560 ATCCCGGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGGTAGTGAAATTTTG 619
QY 427 TTATGTGCCGATGGTGAACCGTGGCTGTCCGGTCTGCTGCTGCTGCTGCTGCTGCTG 486
DB 427 |||||
DB 620 TTATGTGCCGATGGTGAACCGTGGCTGTCCGGTCTGCTGCTGCTGCTGCTGCTGCTG 679
QY 487 AGCGGCGCGAGCTGCGGTTACAAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTT 546
DB 487 |||||
DB 680 AGCGGCGCGAGCTGCGGTTACAAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTT 739
QY 547 ACATCATCAGACATTACCCGGGACTTTATTTAGATAGGCGCGTATGCGCGGCTGTGGGG 606
DB 547 |||||
DB 740 ACATCATCAGACATTACCCGGGACTTTATTTAGATAGGCGCGTATGCGCGGCTGTGGGG 799
```

```
607 CGACGTTCCCGCTCGGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTACCG 666
800 CGACGTTCCCGCTCGGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTACCG 859

667 GCGTACCGCTTGTAATAA 684
860 GCGTACCGCTTGTAATAA 877

RESULT 8
LOCUS      ECOUBICA
DEFINITION Escherichia coli chorismate lyase (ubiC), 4-hydroxybenzoate octaprenyl transferase (ubiA) genes, complete cds, and sn-glycerol-3-phosphate acyltransferase (plsB) genes, 3' end.
ACCESSION  M93136
VERSION    M93136.1 GI:148099
KEYWORDS   4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plsB gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC gene.
SOURCE     Escherichia coli W3110
ORGANISM   Escherichia coli W3110
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE  1 (bases 1 to 2000)
AUTHORS    Lightner,V.A., Bell,R.M. and Modrich,P.
TITLE      The DNA sequences encoding plsB and dgk loci of Escherichia coli J. Biol. Chem. 258 (18), 10856-10861 (1983)
JOURNAL    6309817
PUBMED
REFERENCE  2 (bases 1 to 2000)
AUTHORS    Nichols,B.P. and Green,J.M.
TITLE      Cloning and Sequencing of Escherichia coli ubiC and purification of chorismate lyase
JOURNAL    Unpublished (1992)
COMMENT    Original source text: Escherichia coli (sub_strain W3110, strain K-12) (library: Kohare lambda-1F8) DNA.
FEATURES
Location/Qualifiers
1..2000
/organism='Escherichia coli W3110'
/mol_type='genomic DNA'
/strain='K-12'
/sub_strain='W3110'
/db_xref='taxon:316407'
/maps='91.8 min'
/tissue.lib='Kohare lambda-1F8'
380..877
/gene='ubiC'
380..877
/gene='ubiC'
/function='enzymatic chorismate --> p-hydroxybenzoate + pyruvate'
/experiment='experimental evidence, no additional details recorded'
/citation=[2]
/codon_start=1
/transl_table=11
/product='chorismate lyase'
/protein_id='AAA24711.1'
/db_xref='GI:148100'
/translacion='MSHPALTQLRALRYCKEIPALDPQLLDWLLLEDSMTKRFEQOQK TVSVTMIREGFVEQNEIPEELPLPKESRYLWREILLCADGPFMLAGRTVVPVSTLSG PELALQKLKTPLGRLYFTSSTLTDRDFEIGRDAGLWGRSRRLSLSGKPELLLTFLPL ASPLY'
890..1762
/gene='ubiA'
890..1762
/gene='ubiA'
/function='enzymatic - 3-octaprenyl-4-hydroxybenzoate synthesis'
/experiment='experimental evidence, no additional details recorded'
/citation=[2]
```

gene complete cds, chorismate lyase (ubiC) gene complete cds, sn-glycerol-3-phosphate acyltransferase (plbB) gene, 3' end.		ACCESSION M93413.1 VERSION M93413.1 KEYWORDS 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC gene.
SOURCE Escherichia coli W3110		ORGANISM Escherichia coli W3110
REFERENCE 1 (bases 1 to 2000)		AUTHORS Lightner, V.A., Bell, R.M. and Modrich, P.
TITLE The DNA sequences encoding plbB and dgk loci of Escherichia coli		JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)
PUBMED 6309817		
REFERENCE 2 (bases 1 to 2000)		AUTHORS Nichols, B.P. and Green, J.M.
TITLE Cloning and sequencing of Escherichia coli ubiC and purification of chorismate lyase		JOURNAL J. Bacteriol. 174 (16), 5309-5316 (1992)
PUBMED 1644758		
COMMENT Original source text: Escherichia coli (sub_strain W3110, strain K-12) (library: Kohara lambda) DNA.		
FEATURES Location/Qualifiers		
1..2000		
/organism="Escherichia coli W3110"		
/mol_type="genomic DNA"		
/strain="K-12"		
/sub_strain="W3110"		
/db_xref="taxon:316407"		
/map="91.8 min"		
/tissue_lib="Kohara lambda"		
380..877		gene
/gene="ubiC"		
380..877		CDS
/genes="ubiC"		
/functions="enzymatic chorismate --> p-hydroxybenzoate + pyruvate"		
/experiment="experimental evidence, no additional details recorded"		
/notes="N-terminal amino acid sequence of chorismate lyase :NH2-SHPALTQLRALRYCEIPAL-"		
/citation=[2]		
/codon_start=1		
/transl_table=11		
/product="chorismate lyase"		
/protein_id="AAA24716.1"		
/db_xref="GI:148107"		
/translation="MSPHALTQLRALRYCKEIPALDPQLLDMLLEDSTMTRFBOQKG TVSVTMIREGFVEQNEIPBELPLLPKESRYWLREILLCADGEPWLAGRTVYVPVSTLSG PELALQKLKGTPLGRLYFTSTLTDRDFIEIGRDAGLWRRRLRSKPLLTLEFLP ASPLY"		
890..1762		gene
/gene="ubiA"		
890..1762		CDS
/genes="ubiA"		
/functions="enzymatic - 3-octaprenyl-4-hydroxybenzoate synthesis"		
/experiment="experimental evidence, no additional details recorded"		
/citation=[2]		
/number=2		
/codon_start=1		
/transl_table=11		
/product="4-hydroxybenzoate-octaprenyl transferase"		
/protein_id="AAA24717.1"		
/db_xref="GI:148108"		
/translation="MWSLTQNKLLAFHRLMRTDKPIGALLLLWPTLWALWVATPGVP QLWILAVFVAGVWLMRAAGCVNDYADRKFDGHVKRTANRPLPSGAVTEKARALFVV LVLSFLVLNTMTILLSIAALAWYFPMKRYTHLPQVVLGAAGFWSIPMAFAA VSEVPLSCWLMFLANIUMAVDTQYAMVDRDDDKIKSTAILFGQYDKLIIGIL QIGVLAIAIIGELNGWGYWSILVAGALFVYQOKLIANREREAFCFAPMNNYVG		

/codon_start=1		/transl_table=11		/product="4-hydroxybenzoate-octaprenyl transferase"		/protein_id="AAA24712.1"		/db_xref="GI:148101"		/translation="MWSLTQNKLLAFHRLMRTDKPIGALLLLWPTLWALWATPGVP OLWILAVFVAGVWLMRAAGCVNDYADRKFDGHVKRTANRPLPSGAVTEKARALFVV LVLSFLVLNTMTILLSIAALAWYFPMKRYTHLPQVVLGAAGFWSIPMAFAA VSEVPLSCWLMFLANIUMAVDTQYAMVDRDDDKIKSTAILFGQYDKLIIGIL QIGVLAIAIIGELNGWGYWSILVAGALFVYQOKLIANREREAFCFAPMNNYVG LVFLGLAMSYWHF"		complement (1917..2000)		/gene="plbB"		complement (1917..2000)		/gene="plbB"		/EC_number="2.3.1.15"		/function="lysophosphatidic acid synthesis"		/experiment="experimental evidence, no additional details recorded"		/citation=[1]		/codon_start=1		/transl_table=11		/product="sn-glycerol-3-phosphate acyltransferase"		/protein_id="AAA24713.1"		/db_xref="GI:148102"		/translation="MKVYQLLAELITSDVRLTIESATQGEQ"																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Query Match		72.8%;		Score 498;		DB 15;		Length 2000;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											

gene
CDS
LVLFLGLAMSYWHP"
complement (1917. .2000)
/gene="pleB"
complement (1917. .2000)
/gene="pleB"
/BC_number="2.3.1.15"
/function="lysophosphatidic acid synthesis"
/experiment="experimental evidence, no additional details
recorded"
/citation=[1]
/codon_start=1
/transl_table=11
/product="sn-glycerol-3-phosphate acyltransferase"
/protein_id="AA24718.1"
/db_xref="GI:148109"
/translation="MKVYQLLAELITSDVRLTIESATQEGE"
ORIGIN
Query Match 72.8%; Score 498; DB 15; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTACACCCCGCTTAACGCACTGCGTGGCTGCTGAGATTCCATGACAAACGTTTGA 246
Db |||||
QY 247 CTGGATCCGCACTGCTGCACTGGCTGTTGCTGGAGATTCCATGACAAACGTTTGA 306
Db |||||
QY 440 CTGGATCCGCACTGCTGCACTGGCTGTTGCTGGAGATTCCATGACAAACGTTTGA 499
Db |||||
QY 307 CAGCAGGGAAAAACGTAAGCGTAGCGATATCCGCGAAGGGTTTCTCGAGCAGAAATGA 366
Db |||||
QY 500 CAGCAGGGAAAAACGTAAGCGTAGCGATATCCGCGAAGGGTTTCTCGAGCAGAAATGA 559
QY 367 ATCCCGGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGTTAGTGAAATTTG 426
Db |||||
QY 560 ATCCCGGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGTTAGTGAAATTTG 619
QY 427 TTATGTGCGGATGTTGAACCGTGGCTTGGCGGTCGTACCGTCTCTGTCTCAACGTTA 486
Db |||||
QY 620 TTATGTGCGGATGTTGAACCGTGGCTTGGCGGTCGTACCGTCTCTGTCTCAACGTTA 679
QY 487 AGCGGCGCGAGTGGCTTTACAAAAATTTGGGTAAACCCGCTAGGACGCTATCTGTT 546
Db |||||
QY 680 AGCGGCGCGAGTGGCTTTACAAAAATTTGGGTAAACCCGCTAGGACGCTATCTGTT 739
QY 547 ACATCATGCACATTACCCGGGACTTTATGAGATAGGCGGTGATCCCGGCTGTGGGG 606
Db |||||
QY 740 ACATCATGCACATTACCCGGGACTTTATGAGATAGGCGGTGATCCCGGCTGTGGGG 799
QY 607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGTAACAGAACTGTTTTTACCG 666
Db |||||
QY 800 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGTAACAGAACTGTTTTTACCG 859
QY 667 GCGTACCGCTGTACTAA 684
Db |||||
QY 860 GCGTACCGCTGTACTAA 877

RESULT 10
ECUBIAC
LOCUS
DEFINITION
E.coli ubiC and ubiA genes for chorismate lyase and
4-hydroxybenzoate octaprenyltransferase.
ACCESSION
X57434
VERSION
X57434.1 GI:43233
KEYWORDS
4-hydroxybenzoate octaprenyl transferase; 4-hydroxybenzoate
synthetase; chorismate lyase.
SOURCE
Escherichia coli
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1

AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
Nishimura, K., Nakahigashi, K. and Inokuchi, H.
Location of the ubiA gene on the physical map of Escherichia coli
J. Bacteriol. 174 (17), 5762 (1992)
1512213
2 (bases 1 to 2034)
Nishimura, K.
Direct Submission
Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto
University, Sakyo-ku, Kyoto 606, Japan
FEATURES
Location/Qualifiers
source
1. .2034
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12 W3110"
/db_xref="taxon:562"
/map="91.7 min"
/clone="634"
/clone_lib="Kohara library"
-35_signal
137.142
-10_signal
159.164
gene
291.899
CDS
291.899
/gene="ubiC"
/gene="ubiC"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate synthetase"
/protein_id="CAA0681.1"
/db_xref="GI:43234"
/db_xref="GOA:P26602"
/db_xref="UniProtKB/Swiss-Prot:P26602"
/translation="MRLRFCCVLDHLICFTSPVNTFLRYNAFTLCNGBFGMSHPALT
QLRLAYKEIIPALDQDLWLLLEDSTKRPEQOGKTVSTMIRGFEVQNEIPEL
PLPKESRYLREILLCADGEPWLAGRTVVPVSTLGSPELAKLQKTKPLGRILFTSS
TLTRDFIEIGRDAGLWRRSLRSLSGKLLLELFLPASPLY"
912.1784
gene
912.1784
CDS
912.1784
/gene="ubiA"
/gene="ubiA"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate octaprenyltransferase"
/protein_id="CAA0682.1"
/db_xref="GI:43235"
/db_xref="GOA:P26601"
/db_xref="UniProtKB/Swiss-Prot:P26601"
/translation="NEWSLTQNKLLAFHRLMTDRPIGALLLLWPTLWALWATPGVP
QWLILAVFVAGVWLMRAAGCVVDYADRFKDHVXRTANRPLPSGAVTKEKALFV
LVLIISFLVLTNTMTILLSTAAALAWYPPMKRYTHLPQVVLGAAGFWSIPMAFAA
VSESVPKSWLFLANILWAVDTQYAMVDRDDVKIGIKSTAILFGQYDKLIIIGIL
QIGVLALMAIIGELNGLGWYYSILVAGALFVYQKLIANRERBACPKAFMNNVVG
LVLFLGLAMSYWHP"
ORIGIN
Query Match 72.8%; Score 498; DB 15; Length 2034;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTACACCCCGCTTAACGCACTGCGTGGCTGCTGATTTGTAAGAGATCCCTGCC 246
Db |||||
QY 402 ATGTACACCCCGCTTAACGCACTGCGTGGCTGCTGATTTGTAAGAGATCCCTGCC 461
Db |||||
QY 247 CTGGATCCGCACTGCTGCACTGGCTGTTGCTGGAGATTCCATGACAAACGTTTGA 306
Db |||||
QY 462 CTGGATCCGCACTGCTGCACTGGCTGTTGCTGGAGATTCCATGACAAACGTTTGA 521
Db |||||
QY 307 CAGCAGGGAAAAACGTAAGCGTAGCGATGATCCGCGAAGGGTTTCTCGAGCAGAAATGA 366
Db |||||
QY 522 CAGCAGGGAAAAACGTAAGCGTAGCGATGATCCGCGAAGGGTTTCTCGAGCAGAAATGA 591
Db |||||
QY 367 ATCCCGGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGTTAGTGAAATTTG 426
Db |||||
QY 582 ATCCCGGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTACTGTTAGTGAAATTTG 641

AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
Nishimura, K., Nakahigashi, K. and Inokuchi, H.
Location of the ubiA gene on the physical map of Escherichia coli
J. Bacteriol. 174 (17), 5762 (1992)
1512213
2 (bases 1 to 2034)
Nishimura, K.
Direct Submission
Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto
University, Sakyo-ku, Kyoto 606, Japan
FEATURES
Location/Qualifiers
source
1. .2034
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12 W3110"
/db_xref="taxon:562"
/map="91.7 min"
/clone="634"
/clone_lib="Kohara library"
-35_signal
137.142
-10_signal
159.164
gene
291.899
CDS
291.899
/gene="ubiC"
/gene="ubiC"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate synthetase"
/protein_id="CAA0681.1"
/db_xref="GI:43234"
/db_xref="GOA:P26602"
/db_xref="UniProtKB/Swiss-Prot:P26602"
/translation="MRLRFCCVLDHLICFTSPVNTFLRYNAFTLCNGBFGMSHPALT
QLRLAYKEIIPALDQDLWLLLEDSTKRPEQOGKTVSTMIRGFEVQNEIPEL
PLPKESRYLREILLCADGEPWLAGRTVVPVSTLGSPELAKLQKTKPLGRILFTSS
TLTRDFIEIGRDAGLWRRSLRSLSGKLLLELFLPASPLY"
912.1784
gene
912.1784
CDS
912.1784
/gene="ubiA"
/gene="ubiA"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate octaprenyltransferase"
/protein_id="CAA0682.1"
/db_xref="GI:43235"
/db_xref="GOA:P26601"
/db_xref="UniProtKB/Swiss-Prot:P26601"
/translation="NEWSLTQNKLLAFHRLMTDRPIGALLLLWPTLWALWATPGVP
QWLILAVFVAGVWLMRAAGCVVDYADRFKDHVXRTANRPLPSGAVTKEKALFV
LVLIISFLVLTNTMTILLSTAAALAWYPPMKRYTHLPQVVLGAAGFWSIPMAFAA
VSESVPKSWLFLANILWAVDTQYAMVDRDDVKIGIKSTAILFGQYDKLIIIGIL
QIGVLALMAIIGELNGLGWYYSILVAGALFVYQKLIANRERBACPKAFMNNVVG
LVLFLGLAMSYWHP"

QY 427 TTATGTGCGCATGGTAACCGGTGGCTTCCGGGTCTACCGTGGTTCCTGTGTCAACGTTA 486
Db 642 TTATGTGCGCATGGTAACCGGTGGCTTCCGGGTCTACCGTGGTTCCTGTGTCAACGTTA 701
QY 487 AGCGGGCGGAGCTGGCTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 546
Db 702 AGCGGGCGGAGCTGGCTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 761
QY 547 ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCGCGTGTGATGCGGGCTGTGGGG 606
Db 762 ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCGCGTGTGATGCGGGCTGTGGGG 821
QY 607 CGACGTTCCCGCTCGATTAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 666
Db 822 CGACGTTCCCGCTCGATTAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 881
QY 667 GCGTCACCGTTGTACTAA 684
Db 882 GCGTCACCGTTGTACTAA 899

RESULT 11
ECUBI
LOCUS 2348 bp DNA linear BCT 09-SEP-2004
DEFINITION E.coli genes ubiC and ubiA.
ACCESSION X66619.1
VERSION X66619.1 GI:43230
KEYWORDS 4-hydroxybenzoate-octaprenyltransferase; chorismate-pyruvate lyase; ubiA gene; ubiC gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 2348)
AUTHORS Siebert, M., Rechold, A., Melzer, M., May, U., Berger, U., Schroder, G., Schroder, J., Severin, K. and Heide, L.
TITLE Ubiquinone biosynthesis. Cloning of the genes coding for chorismate pyruvate-lyase and 4-hydroxybenzoate octaprenyl transferase from Escherichia coli
JOURNAL FEBS Lett. 307 (3), 347-350 (1992)
PUBMED 1644192
REFERENCE 2 (bases 1 to 2348)
AUTHORS Heide, L.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-1992) L. Heide, Inst of Pharmaceutical Biology, Schaenzlestrasse 1, 7800 Freiburg, FRG
COMMENT On Sep 24, 2004 this sequence version replaced gi:253104.
See also K00127, M93136 & M93413.
FEATURES
source
1. .2348
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="MC4100"
/db_xref="taxon:562"
/map="91.8 mins"
/clone="pALMUI"
/clone_lib="Sau3A partial digested DNA in pBluescript"
558. .563
570. .1067
/gene="ubiC"
570. .1067
/gene="ubiC"
/codon_start=1
/transl_table=11
/product="chorismate-pyruvate lyase"
/protein_id="CAA47181.1"
/db_xref="GI:43231"
/db_xref="GOA:P26602"
/db_xref="UniProtKB/Swiss-Prot:P26602"
/translation="MSHPALQRLARYCKEIPALDPQLLDLWLLLEDSMTKRFRQOQK TVSVTMIREGVQNEIPEELPLLPKESRYWLREILLCADGEPWLAGRTVPVSTLSG PELALQKLGKTPLGRLYFTSITLTRDFEIGRDAGLWGRSRLRLSGKPLLLTELFLP

ASPLY"
1068. .1073
1080. .1952
/gene="ubiA"
1080. .1952
/gene="ubiA"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate-octaprenyl transferase"
/protein_id="CAA47182.1"
/db_xref="GI:43232"
/db_xref="GOA:P26601"
/db_xref="UniProtKB/Swiss-Prot:P26601"
/translation="MWSLQNKLLAFHRLNRTDKPIGALLLLWPTLWALWVATPGVP QLTILAVFVAGVWLMRAAGCVVNDYADKKFDGHVKRYTHANRPLPSGAVTEKGRALFV LVLISFLVLTNTMTITLLSIALALAWVYPMKRYTHLPQVVLGAAFGWISIPWAFAA VSESVPLSCWLMFLANILWAVDTQYAMVDRDDDKIKISTAILFQYDKLIIGIL QIGVLAALMAIIGELNGWGYWSILVAGALFVYQOKLIANRERACFKAFMNNYVG LVLPLGLMSYWHF"
1254. .1322
misc_binding
/gene="ubiA"
/bound_moiety="polyprenyl"
1965. .1989
terminator
repeat_region 1994. .2096
/note="palindrome"
ORIGIN
Query Match 72.8%; Score 498; DB 15; Length 2348;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 246
Db 570 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 629
QY 247 CTGGATCCGCAACTGCTGACTGCTGTTGCTGGAGGATTCATGACAAAAAGTTTGA 306
Db 630 CTGGATCCGCAACTGCTGACTGCTGTTGCTGGAGGATTCATGACAAAAAGTTTGA 689
QY 307 CAGCAGGAAAAACCGTAAAGCGTCACGATGATCCCGAAGGGTTTGTGACGAGAAATGA 366
Db 690 CAGCAGGAAAAACCGTAAAGCGTCACGATGATCCCGAAGGGTTTGTGACGAGAAATGA 749
QY 367 ATCCCGGAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 426
Db 750 ATCCCGGAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 809
QY 427 TTATGTGCCGATGGTAACCGTGGCTTCCCGCTGCTACCGTCTGTTCTGTGTCAACGTTA 486
Db 810 TTATGTGCCGATGGTAACCGTGGCTTCCCGCTGCTACCGTCTGTTCTGTGTCAACGTTA 869
QY 487 AGCGGGCGGAGCTGGGCTTACAAAATTGGGTAAACCGCGTGTAGGACGCTATCTGTTTC 546
Db 870 AGCGGGCGGAGCTGGGCTTACAAAATTGGGTAAACCGCGTGTAGGACGCTATCTGTTTC 929
QY 547 ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCGCGTGTAGGACGCTGTGGGG 606
Db 930 ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCGCGTGTAGGACGCTGTGGGG 989
QY 607 CGAGGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 666
Db 990 CGAGGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACGAACTGTTTTACCG 1049
QY 667 GCGTCACCGTTGTACTAA 684
Db 1050 GCGTCACCGTTGTACTAA 1067
RESULT 12
AR144789
LOCUS
DEFINITION Sequence 25 from patent US 6210937.
ACCESSION AR144789

AR144789 6641 bp DNA linear PAT 08-AUG-2001
Sequence 25 from patent US 6210937.
AR144789

RESULT 13
U00096_42
WFCOMMENT
Sequence split into 47 fragments LOCUS U00096 Accession U00096
Fragment Name Begin End
U00096_00 1 110000
U00096_01 100001 210000
U00096_02 200001 310000
U00096_03 300001 410000
U00096_04 400001 510000
U00096_05 500001 610000
U00096_06 600001 710000
U00096_07 700001 810000
U00096_08 800001 910000
U00096_09 900001 1010000

U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000
U00096_14	1400001	1510000
U00096_15	1500001	1610000
U00096_16	1600001	1710000
U00096_17	1700001	1810000
U00096_18	1800001	1910000
U00096_19	1900001	2010000
U00096_20	2000001	2110000
U00096_21	2100001	2210000
U00096_22	2200001	2310000
U00096_23	2300001	2410000
U00096_24	2400001	2510000
U00096_25	2500001	2610000
U00096_26	2600001	2710000
U00096_27	2700001	2810000
U00096_28	2800001	2910000
U00096_29	2900001	3010000
U00096_30	3000001	3110000
U00096_31	3100001	3210000
U00096_32	3200001	3310000
U00096_33	3300001	3410000
U00096_34	3400001	3510000
U00096_35	3500001	3610000
U00096_36	3600001	3710000
U00096_37	3700001	3810000
U00096_38	3800001	3910000
U00096_39	3900001	4010000
U00096_40	4000001	4110000
U00096_41	4100001	4210000
U00096_42	4200001	4310000
U00096_43	4300001	4410000
U00096_44	4400001	4510000
U00096_45	4500001	4610000
U00096_46	4600001	4639675
Continuation (43 of 47) of U00096 from base 4200001 (U00096 Escherichia coli K-12)		
Query Match 72.8%; Score 498; DB 15; Length 110000;		
Best Local Similarity 100.0%; Pred. No. 3e-143;		
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	187	ATGTCACACCCCGGTTTAACGCAACTGCGTGCCTGCGCTATTGTAAAGAGATCCCTGCC 246
Db	50529	ATGTCACACCCCGGTTTAACGCAACTGCGTGCCTGCGCTATTGTAAAGAGATCCCTGCC 50588
Qy	247	CTGGATCGCAACTGCTCGACTGCTGTTGTGGAGGATTCATGACAAAAACGTTTGGAA 306
Db	50589	CTGGATCGCAACTGCTCGACTGCTGTTGTGGAGGATTCATGACAAAAACGTTTGGAA 50648
Qy	307	CAGCAGGAAAAACGGTAAAGCGGTGACGATGATCCCGAGGGTTTGTCCGACGAGAATGAA 366
Db	50649	CAGCAGGAAAAACGGTAAAGCGGTGACGATGATCCCGAGGGTTTGTCCGACGAGAATGAA 50708
Qy	367	ATCCCCGAAGAACTGCCCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426
Db	50709	ATCCCCGAGAAGTGCCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 50768
Qy	427	TTATGTGCCGATGGTGAACCGTGGCTTCGCCGTCCGTACCGTGGTTCCTGTGTCAACGTTA 486
Db	50769	TTATGTGCCGATGGTGAACCGTGGCTTCGCCGTCCGTACCGTGGTTCCTGTGTCAACGTTA 50828
Qy	487	AGCGGCCGGAGCTGGCGTTACAAAAATTGGTAAACGCCGTTAGGACGCTATCTGTTTC 546
Db	50829	AGCGGCCGGAGCTGGCGTTACAAAAATTGGTAAACGCCGTTAGGACGCTATCTGTTTC 50888
Qy	547	ACATCATCGACATTAAACCCGGGACTTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGGG 606
Db	50889	ACATCATCGACATTAAACCCGGGACTTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGGG 50948
Qy	607	CGACGTTCCCGCCTCGCATTAAGCGGTTAAACCGCTGTTGCTTAACAGAACTGTTTATCCG 666

misc_feature
1. .1737
/notes="corresponds to M21516; ECOKATGA (1225. .2805) "
/gene="katG"
CDS
1. .1257
/genes="katG"
/EC_number="1.11.1.6"
/standard_name="catalase HPI; catalase-peroxide; catalase"
/notes="CG Site No. 14983"
/codon_start=1
/transl_table=11
/product="catalase hydroperoxidase I"
/protein_id="AAC43048.1"
/db_xref="GI:396289"
/translation="AITSGLEVWVTTQTQWNSYFENLTKYEWVQTRSPAGAIOPEA
VPAEIIIPDPFKSRKPKTMLVTDLTRDFEFKISRRLNDPOAFNEAFARXWFK
LTHRDMGPKSRVIGPEVKEDLIWQDPLPQPIYNPEQDIIDLKFAIDSGLSVSELY
SVNVAASFTFRGDKRGKANGARKLAMPORDWDVNAARALPVLKIKESGKSASLA
DIIVAGVVGVSKAASAGLSHVFPAPGRVDARQDQDIEMFELLEPIADGFRNYRA
RLDVSTESLLIDKAQQLTAPEMTALVGMRLGANFDGSKNGVFTDRVGVLNDF
FVNLLDMRYEMKATDSKELFEGRDRETGEVKTASRADLVFGSNSVLRAVAEYVASS
DAHEKFDKDFVAWVKVMDLRFDLL"
misc_feature
1. .161
/genes="katG"
/notes="corresponds to lambda clone EC18-126"
misc_feature
1. .93
/genes="katG"
/note="93 bp overlap with end of Li9201 (ECOWU87) "
misc_feature
156. .15760
/notes="corresponds to lambda clone EC14-54"
misc_difference
938
/genes="katG"
/note="G in M21516; C here"
1266. .1305
/notes="putative"
1317. .2255
/note="similar to Desulfurolobus ambivalens hypoth. 28.3
kDa protein in sor 3' region"
/codon_start=1
/transl_table=11
/label=ORF_O312
/protein_id="AAC43049.1"
/db_xref="GI:409786"
/translation="MGERKLHTGSLMSAAGKNPLAISGLVLTILWISYWMFKQVT
VYIGAFDFALRCIFGALVFIVLLLRGRGMPTFPKTYTLAIALQTQCMVGLAQWAL
SVSGAGKVAIILSYTWTFMWVVIIFALFLGERLGRGOYFALLIAFGLFLVLPQWLDPS
SNKSAAILLSGVSGASAIKRIYARHPRVLDLLSLTSWQMLYALVMSVVALVPQ
REIDQPTVFWALAYSAILATALWSMLFVLKLNLPASIASLSTILAVPVCGLVFSWWL
LGENPGVGGSGIVLIVLALALVSRKKKEAVSVKRI"
misc_difference
1429. .1432
/note="TTT in M21516; TTTT here"
misc_difference
1720
/note="C in M21516; G here"
CDS
complement(2282. .2899)
/codon_start=1
/transl_table=11
/label=ORF_F205
/protein_id="AAC43050.1"
/db_xref="GI:396291"
/translation="WKASLALLSLTATFTSHLSKSPAVPPTVVQIQTANNLAIADGAR
QQIGSTLYDPAYQLTYPGGDVQPERGVCSDVIRALRSQKVDLQKLVDHDMKNFA
EYQWKLRKRPDSNIDHRVPNLETWFSRHDKTRPTSKNPSDYQAGDQIVSWRLDGLA
HIGVYDGFARDGTPLVTIHNICAGAEEDVLFNRMVGHYRFVK"
promoter
complement(2886. .2915)
/note="promoter matrix score of 42; putative; within ORF
f205, which would suggest alternate start codon"
misc_structure
2912. .3053
/note="predicted bend of 81 degrees"
repeat_region
3031. .3064
/standard_name="REP; repetitive extragenic palindromic
element"
/note="contains 1 REP sequence"
misc_structure
3090. .3219

gene
complement(3174. .4316)
/gene="gldA"
CDS
complement(3174. .4316)
/gene="gldA"
/note="similar to Bacillus stearothermophilus glycerol
dehydrogenase"
/codon_start=1
/transl_table=11
/label=ORF_F380
/product="Glycerol dehydrogenase"
/protein_id="AAC43051.1"
/db_xref="GI:396292"
/translation="NPHLALLISKGAIMDRIIIOSPKYIQGADVIRNLGRLKPLXER
WLVGDKFVLGPAQSTVKEKSPDKAGLVVIAIEPFGECSCNIEDRLRIARTACOGAIL
GIGGKTLDTAKALAHFMGVVAIAPTIASTDAPCSAHSVITYTDEGEFDRYLLPNP
NNVITKIVAGAPARLLAAGIGDALATWFEARACSRGATWAGCKTQAALALAEI
CYNTHLEGEKAMLAEOHVVTPALERVIENANTYLSGVGFESGGLAAAHAVHGLTAI
PDHHYHGEKVAFGTLTQVLENAPVEIETVAALSHAVGLPITITLAQDLIKEDVPAK
MRIVAAACAEGEITHNMPGGATPDQVYAALLVADQYQGRFLQEME"
misc_structure
4022. .4156
/note="predicted bend of 75 degrees"
CDS
complement(4288. .4950)
/note="similar to Bacillus subtilis hypoth. 20 kDa
protein, in tsr 3' region"
/codon_start=1
/transl_table=11
/label=ORF_F220
/protein_id="AAC43052.1"
/db_xref="GI:396293"
/translation="MELYLDTANVAEVLRLARIFPIAGVTTNPSIIAASKESIWEVL
RLQKAIGDEGILFAOTMSRDAQMVEEAKRLDAIPGVVKIPVTSEGLAAIKILKE
GITTLGTAVYSAOGLLAALAGAKYVAVVNEVDAGGDIQRTVQKLTLLNHAPES
MYLAASFYTPROALDCLLAGCESITLPLDVAQOQMLNTPAVESAIEFEHDWNAAFGTT
HL"
promoter
complement(4922. .4949)
/note="promoter matrix score of 47; putative; within ORF
f220, which would suggest alternate start codon"
CDS
complement(4962. .7097)
/note="similar to phosphotransferase system enzyme I"
/codon_start=1
/transl_table=11
/label=ORF_F711
/protein_id="AAC43053.1"
/db_xref="GI:409787"
/translation="MCSSSAGGILTPISLDDLNALGNLPAKGVDAEQSALENGLTIV
LKNIFRLDDSGATSAILEAHSRSLAGDTSLEHLLAGVSAGLSACAIIVASANHFCE
EFSRSSYLOERALDVRDVCQLLQIYGEORFPAPKLTQPAICMADELTPSQFLE
LDKNHLGILLKSGGTTSHTVILARFMTLVGVVDIDALTTPWQQOTIYIDNAGAI
VEPGEAVARYQOEAREVQDALREQORVMTLQOARTADGIRIBIANIASHVEAQAAPF
NGAEGVGLFRTEMLYMDRTSAPGESELYNIFCOALESANGRSIIIVRTMISGDKPVDY
LNIPEANPFILGYRAVRIYEYASLFTQLRSILRASAGSLKIMPMISSMBEILWY
KEKLAERQQLRNEHIPFEKIQGLMIEVPSVMFTIDQCCBEIDFFSIGNDLTQYL
LAVDRDNKRVTHNINSLPAFLDALYAVQVHRQGWIGLCELGAKGSLVPLLVGL
GLDELSMGAISPAPAKMAQDSECRLLNQAMACRTSLSEVHLLAQFRMTQDDAP
LVTAEICITLESWRSKEEVLKGMTNLLLAGRCYPRKLEADLWAREAVFTGLGFSF
AIPHSKSHIEQSTISVARLOAPVWGDDEAQFIIMLTINKHAAGQDHMRIFSLRLAR
IMHEEFNALVNAASADAIALSLQHELEL"
CDS
7099. .7428
/codon_start=1
/transl_table=11
/label=ORF_O109
/protein_id="AAC43054.1"
/db_xref="GI:409788"
/translation="MRARIICGFRILVSETSGSSSSDLTTSAGSQMGNSSRSWLKRR
WCACSSAPEISNWQLSPASVSPISARALLALPCGCAGYATQPMVKKYTYAFQRGL
PDAHVR"
promoter
complement(7108. .7136)
/note="promoter matrix score of 46; putative"
misc_feature
7200. .25839
/note="corresponds to lambda clone EC21-52"
terminator
7578. .7699
/note="putative"

CDS	7771..8850	/note="similar to phosphotransferase system enzyme II"
	/codon start=1	
	/transl table=11	
	/label=ORF_0359	
	/protein_id="AAC43055.1"	
	/db_xref="GI:396296"	
	/translation="MNELVQILKNTROHLMGTVMSPHFVSSGGTLLAVSMVLYGKGA	
	VPDAVADPNKKLPIGVAGLTLMVPFLAAVIGYSIAERSALAPCAIGAWVGNSTGAG	
	PFGLAGLIGCVIVHYHLKIPVHKVLSVMPFIIPIVGTLTITAGIMWMLGEPVGA	
	LTNSLTQWLQMGQOQSIYMLAVINGMLMLAFDMGGPVNKVAYAFMLICVAQGVYTVVAI	
	AAVGICIPPLGMGLATLIGRKNFSABERETGKAALVMGCVGTGTEGAIPFAAADPLRVI	
Query Match	72.8%; Score 498; DB 15; Length 176195;	
Best Local Similarity	100.0%; Pred. No. 3.4e-143;	
Matches 498; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	187 ATGTCACACCCCGGTTTAACGCAACTGCGTGCCTATTTGTAAGAGATCCCTGC 246	
Db	117748 ATGTCACACCCCGGTTTAACGCAACTGCGTGCCTATTTGTAAGAGATCCCTGC 117807	
Qy	247 CTGGATCCGCAACTGCTCGACTGGCTGTTCTGGAGGATTCATGACAAAAAGTTTGAA 306	
Db	117808 CTGGATCCGCAACTGCTCGACTGGCTGTTCTGGAGGATTCATGACAAAAAGTTTGAA 117867	
Qy	307 CAGCAGGGAAAAACGGTAAGCGGTGACGATGATCCCGGAAGGGTTTGTGCGAGCAGAAATGAA 366	
Db	117868 CAGCAGGGAAAAACGGTAAGCGGTGACGATGATCCCGGAAGGGTTTGTGCGAGCAGAAATGAA 117927	
Qy	367 ATCCCCGAAGAACTGCCCGTGTGCGGAAAGAGTCTCGTTACTGTGTACGTGAAATTTTG 426	
Db	117928 ATCCCCGAAGAACTGCCCGTGTGCGGAAAGAGTCTCGTTACTGTGTACGTGAAATTTTG 117987	
Qy	427 TTATGTGCCGATGGTGAACCGTGGCTTCGCGGTCTGACCGTGGTCTGTGTGTCAACGTGA 486	
Db	117988 TTATGTGCCGATGGTGAACCGTGGCTTCGCGGTCTGACCGTGGTCTGTGTGTCAACGTGA 118047	
Qy	487 AGCGGCGCGAGCTGGCGTTACAAAAATGGTAAAAACCGCGTTAGGACGCTATCTGTTC 546	
Db	118048 AGCGGCGCGAGCTGGCGTTACAAAAATGGGTAAAAACCGCGTTAGGACGCTATCTGTTC 118107	
Qy	547 ACATCATCGACATTAAACCCGGACATTATTGAGATAGCGCGTATGCGGGCTGTGGGGG 606	
Db	118108 ACATCATCGACATTAAACCCGGACATTATTGAGATAGCGCGTATGCGGGCTGTGGGGG 118167	
Qy	607 CGACGTTCCCGCCTCGCATTAAGCGGTAAACCGCTGTTGTCTAAACAGAACTGTTTTTACCG 666	
Db	118168 CGACGTTCCCGCCTCGCATTAAGCGGTAAACCGCTGTTGTCTAAACAGAACTGTTTTTACCG 118227	
Qy	667 GCGTCACCGTTGTACTAA 684	
Db	118228 GCGTCACCGTTGTACTAA 118245	

Search completed: May 30, 2006, 02:12:44
Job time : 4102 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM nucleic - nucleic search, using sw model
Run on: May 29, 2006, 10:44:22 ; Search time 322 Seconds
(without alignments)
3974.653 Million cell updates/sec

Title: US-10-718-311-7
Perfect score: 684
Sequence: 1 atggcttcctctgtcatttc.....cggcgtaaccgtgtactaa 684
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 1403666 seqs, 935554401 residues
Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /EMC_Celerra_SIDS3/ptodata/2/ina/1_COMB.seq:*
2: /EMC_Celerra_SIDS3/ptodata/2/ina/5_COMB.seq:*
3: /EMC_Celerra_SIDS3/ptodata/2/ina/6A_COMB.seq:*
4: /EMC_Celerra_SIDS3/ptodata/2/ina/6B_COMB.seq:*
5: /EMC_Celerra_SIDS3/ptodata/2/ina/7_COMB.seq:*
6: /EMC_Celerra_SIDS3/ptodata/2/ina/H_COMB.seq:*
7: /EMC_Celerra_SIDS3/ptodata/2/ina/PCTUS_COMB.seq:*
8: /EMC_Celerra_SIDS3/ptodata/2/ina/PP_COMB.seq:*
9: /EMC_Celerra_SIDS3/ptodata/2/ina/RE_COMB.seq:*
10: /EMC_Celerra_SIDS3/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	684	100.0	684 3 US-09-896-866B-16 Sequence 16, Appl
2	684	100.0	684 3 US-09-855-341-7 Sequence 7, Appl
3	498	72.8	854 3 US-09-064-693A-24 Sequence 24, Appl
4	498	72.8	6641 3 US-09-064-693A-25 Sequence 25, Appl
5	495	72.4	495 3 US-09-896-866B-12 Sequence 12, Appl
6	495	72.4	495 3 US-09-855-341-3 Sequence 3, Appl
7	282	41.2	645 3 US-09-489-039A-1234 Sequence 1234, Ap
8	140.8	20.6	204 3 US-09-839-477-3 Sequence 3, Appl
9	140.8	20.6	499 3 US-09-495-797-41 Sequence 41, Appl
10	133.2	19.5	1442 2 US-08-152-483B-8 Sequence 8, Appl
11	130.6	19.1	177 2 US-08-095-726-12 Sequence 12, Appl
12	130.6	19.1	177 2 US-08-096-043-12 Sequence 12, Appl
13	130.6	19.1	177 2 US-08-093-577-8 Sequence 8, Appl
14	130.6	19.1	177 2 US-08-331-004A-6 Sequence 6, Appl
15	130.6	19.1	177 2 US-08-096-623A-12 Sequence 12, Appl
16	130.6	19.1	177 7 PCT-US95-13937A-6 Sequence 6, Appl
17	126.6	18.5	537 3 US-09-543-681A-3122 Sequence 3122, Ap
18	97.2	14.2	149 3 US-09-839-477-5 Sequence 5, Appl
19	77.2	11.3	281 2 US-08-469-202-15 Sequence 15, Appl
20	77.2	11.3	281 2 US-08-469-202-17 Sequence 17, Appl
21	77.2	11.3	281 2 US-08-469-202-19 Sequence 19, Appl
22	77.2	11.3	281 2 US-08-484-434C-15 Sequence 15, Appl
23	77.2	11.3	281 2 US-08-484-434C-18 Sequence 18, Appl

24	77.2	11.3	281	2	US-08-484-434C-22	Sequence 22, Appl
25	77.2	11.3	281	3	US-09-384-361-15	Sequence 15, Appl
26	77.2	11.3	281	3	US-09-384-361-18	Sequence 18, Appl
27	77.2	11.3	281	3	US-09-384-361-22	Sequence 22, Appl
28	75.2	11.0	104	2	US-08-095-726-38	Sequence 38, Appl
29	75.2	11.0	104	2	US-08-096-043-35	Sequence 35, Appl
30	75.2	11.0	104	2	US-08-093-577-31	Sequence 31, Appl
31	75.2	11.0	104	2	US-08-096-623A-43	Sequence 43, Appl
32	74.4	10.9	264	3	US-09-441-340-9	Sequence 9, Appl
33	74.4	10.9	268	3	US-09-186-002-5	Sequence 5, Appl
34	74.4	10.9	279	2	US-08-391-339-9	Sequence 9, Appl
35	74.4	10.9	279	2	US-08-484-274A-9	Sequence 9, Appl
36	74.4	10.9	279	9	US-09-612-404-9	Sequence 9, Appl
37	74.4	10.9	355	2	US-08-090-523-5	Sequence 5, Appl
38	74.4	10.9	355	2	US-08-398-627-5	Sequence 5, Appl
39	74.4	10.9	355	2	US-08-406-858-5	Sequence 5, Appl
40	74.4	10.9	355	2	US-08-476-519-7	Sequence 7, Appl
41	74.4	10.9	355	3	US-08-120-703A-5	Sequence 5, Appl
42	74.4	10.9	355	3	US-08-399-023-5	Sequence 5, Appl
43	74.4	10.9	355	7	PCT-US91-04036-5	Sequence 5, Appl
44	74.4	10.9	355	7	PCT-US94-05275-5	Sequence 5, Appl
45	74.4	10.9	355	7	PCT-US95-09323-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-09-896-866B-16
; Sequence 16, Application US/09896866B
; Patent No. 6642036
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protei
US-09-896-866B-16

Query Match	100.0%;	Score 684;	DB 3;	Length 684;
Best Local Similarity	100.0%;	Pred. No. 1.1e-230;		
Matches 684;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGGCTTCCTCTGTCAATTTCTTCAGCAGCTGTTCACACGACGACGCAATGTTACACAAGCT	60	
Db	1	ATGGCTTCCTCTGTCAATTTCTTCAGCAGCTGTTCACACGACGACGCAATGTTACACAAGCT	60	
QY	61	AGCATGGTTGACCATCTTCACTGGTCTCTCAATCTTCAGCCACTTTCCCTGTTACAAAGAAG	120	
Db	61	AGCATGGTTGACCATCTTCACTGGTCTCTCAATCTTCAGCCACTTTCCCTGTTACAAAGAAG	120	
QY	121	CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGTCATGCAAGGTG	180	
Db	121	CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGTCATGCAAGGTG	180	
QY	181	TGGCATATGTCACACCCCGGTTAAGCAACTGCGTGCCTATTGTTAAGAGATC	240	
Db	181	TGGCATATGTCACACCCCGGTTAAGCAACTGCGTGCCTATTGTTAAGAGATC	240	

TELEFAX: (208)526-8339
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 854 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-064-693A-24

Query Match 72.8%; Score 498; DB 3; Length 854;
Best Local Similarity 100.0%; Pred. No. 6.2e-165; Mismatches 0; Indels 0; Gaps 0;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTGCGCTATTTGTAAGAGATCCCTGCC 246
DB 321 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTATTTGTAAGAGATCCCTGCC 380
QY 247 CTGGATCCCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 306
DB 381 CTGGATCCCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 440
QY 307 CAGCAGGGAAAAACGTTAGCGTACGATCATCCGGAAGGTTTCTCGAGCAGATGAA 366
DB 441 CAGCAGGGAAAAACGTTAGCGTACGATCATCCGGAAGGTTTCTCGAGCAGATGAA 500
QY 367 ATCCCGAAGAACTGCGCTGCTGCGCAAGAGTCTCTGTTACTGTTAGCTGAAATTTG 426
DB 501 ATCCCGAAGAACTGCGCTGCTGCGCAAGAGTCTCTGTTACTGTTAGCTGAAATTTG 560
QY 427 TTATGTCCGATGGTGAACCGTGGCTGTCGGGTGACCGTCTGTTCTGTGCAACGTTA 486
DB 561 TTATGTCCGATGGTGAACCGTGGCTGTCGGGTGACCGTCTGTTCTGTGCAACGTTA 620
QY 487 AGCGGCGGAGCTGCGTTACAAAATTTGGTTAGAGATGCGGTTAGGAGCTATCTGTT 546
DB 621 AGCGGCGGAGCTGCGTTACAAAATTTGGTTAGAGATGCGGTTAGGAGCTATCTGTT 680
QY 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGATGCCGGGTGTGGGG 606
DB 681 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGATGCCGGGTGTGGGG 740
QY 607 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
DB 741 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800
QY 667 GCGTCACCGTTGACTAA 684
DB 801 GCGTCACCGTTGACTAA 818

RESULT 4
US-09-064-693A-25
Sequence 25, Application US/09064693A
Patent No. 6210937
GENERAL INFORMATION:
APPLICANT: Ward, Thomas E.
TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
ENGINEERED BACTERIA FOR PRODUCTION
OF A SPECIFIC PLASTICS PRECURSOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSER: W. Gary Goodson
ADDRESSEE: INEEL--Lockheed Martin Idaho
ADDRESSEE: Technologies Co.
STREET: P.O. Box 1625
CITY: Idaho Falls
STATE: Idaho
COUNTRY: USA
ZIP: 83415-3810
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: Toshiba Satellite Pro T2150CDS
OPERATING SYSTEM: Windows95

SOFTWARE: Word Perfect 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,693A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: W. Gary Goodson
REGISTRATION NUMBER: 22,387
REFERENCE/DOCKET NUMBER: LIT-PI-296
TELECOMMUNICATION INFORMATION:
TELEPHONE: (208)526-9469
TELEFAX: (208)526-8339
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 6641 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-064-693A-25

Query Match 72.8%; Score 498; DB 3; Length 6641;
Best Local Similarity 100.0%; Pred. No. 2.2e-164; Mismatches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTGCGCTATTTGTAAGAGATCCCTGCC 246
DB 321 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTGCGCTATTTGTAAGAGATCCCTGCC 380
QY 247 CTGGATCCCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 306
DB 381 CTGGATCCCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 440
QY 307 CAGCAGGGAAAAACGTTAGCGTACGATCATCCGGAAGGTTTCTCGAGCAGATGAA 366
DB 441 CAGCAGGGAAAAACGTTAGCGTACGATCATCCGGAAGGTTTCTCGAGCAGATGAA 500
QY 367 ATCCCGGAGAACTGCGCTGCTGCGGAAGAGTCTGTTACTGTTAGCTGAAATTTG 426
DB 501 ATCCCGGAGAACTGCGCTGCTGCGGAAGAGTCTGTTACTGTTAGCTGAAATTTG 560
QY 427 TTATGTCCGATGGTGAACCGTGGCTGTCGGGTGACCGTCTGTTCTGTGCAACGTTA 486
DB 561 TTATGTCCGATGGTGAACCGTGGCTGTCGGGTGACCGTCTGTTCTGTGCAACGTTA 620
QY 487 AGCGGCGGAGCTGCGTTACAAAATTTGGTTAGAGATGCGGTTAGGAGCTATCTGTT 546
DB 621 AGCGGCGGAGCTGCGTTACAAAATTTGGTTAGAGATGCGGTTAGGAGCTATCTGTT 680
QY 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGATGCCGGGTGTGGGG 606
DB 681 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGATGCCGGGTGTGGGG 740
QY 607 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
DB 741 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800
QY 667 GCGTCACCGTTGACTAA 684
DB 801 GCGTCACCGTTGACTAA 818

RESULT 5
US-09-896-866B-12
Sequence 12, Application US/09896866B
Patent No. 6642036
GENERAL INFORMATION:
APPLICANT: Flint, Dennis
APPLICANT: Meyer, Knut
APPLICANT: Viitanen, Paul
TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate:

```
; TITLE OF INVENTION: Benzoin Acid Glucosides
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12

Query Match      72.4%; Score 495; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.1e-164;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246
Db 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 60

Qy 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTGA 306
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 307 CAGCAGGGAACAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGACGAGATGAA 366
Db 121 CAGCAGGGAACAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGACGAGATGAA 180

Qy 367 ATCCCGGAAGAACTGCCGCTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 426
Db 181 ATCCCGGAAGAACTGCCGCTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 240

Qy 427 TTATGTCCGATGTGTAACCGGTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 486
Db 241 TTATGTCCGATGTGTAACCGGTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 300

Qy 487 AGCGGGCCGAGCTGGGCTTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTT 546
Db 301 AGCGGGCCGAGCTGGGCTTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTT 360

Qy 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 606
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 420

Qy 607 CGAGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 666
Db 421 CGAGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480

Qy 667 GCGTCACCGTTGTAC 681
Db 481 GCGTCACCGTTGTAC 495
```

```
RESULT 6
US-09-855-341-3
; Sequence 3, Application US/09855341
; Patent No. 6683231
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
```

```
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-09-855-341-3

Query Match      72.4%; Score 495; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.1e-164;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246
Db 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 60

Qy 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTGA 306
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 307 CAGCAGGGAACAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGACGAGATGAA 366
Db 121 CAGCAGGGAACAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGACGAGATGAA 180

Qy 367 ATCCCGGAAGAACTGCCGCTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 426
Db 181 ATCCCGGAAGAACTGCCGCTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 240

Qy 427 TTATGTCCGATGTGTAACCGGTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 486
Db 241 TTATGTCCGATGTGTAACCGGTGCTGCGGAAAGAGTCTGTTACTGTTACGTTGAAATTTG 300

Qy 487 AGCGGGCCGAGCTGGGCTTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTT 546
Db 301 AGCGGGCCGAGCTGGGCTTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTT 360

Qy 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 606
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 420

Qy 607 CGAGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 666
Db 421 CGAGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480

Qy 667 GCGTCACCGTTGTAC 681
Db 481 GCGTCACCGTTGTAC 495
```

```
RESULT 7
US-09-489-039A-1234
; Sequence 1234, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1234
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1234

Query Match      41.2%; Score 282; DB 3; Length 645;
Best Local Similarity 72.9%; Pred. No. 1e-88;
Matches 363; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246
Db 148 ATGTCACATCCTGGCTTACGCGACTGCGTGGCTGCGCTATTTTGGCGGTTATGCCATCC 207
```

```

RESULT 9
US-09-435-797-41
; Sequence 41, Application US/09495797
; Patent No. 6369296
; GENERAL INFORMATION:
; APPLICANT: Ratcliff, Frank G
; APPLICANT: Martin-Hernandez, Ana M
; APPLICANT: Bulcombe, David C
; TITLE OF INVENTION: Viral Vectors
; FILE REFERENCE: Mewburn 43, 047
; CURRENT APPLICATION NUMBER: US/09/495,797
; CURRENT FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: N. benthamiana rubisco
US-09-435-797-41

```

Query Match	20.6%	Score 140.8;	DB 3;	Length 499;
Best Local Similarity	85.3%	Pred. No. 6.3e-39;		
Matches 157; Conservative	0;	Mismatches 27;	Indels 0;	Gaps 0

Qy	1	ATGGCTTCTCTGTCATTTCTTCAGCAGCTGTGCCACACGACGAAGTTTACACAAGCT	60
Db	8	ATGGCTTCTCTGTCATTTCTTCAGCAGCTGTGCCACCGCAGCAATGTTCGCTCAAGCT	67
Qy	61	AGCATGGTTGCACCTTTTCACCTGGTCTCAAAATCTTCAGCGCACCTTTCCCTGTTTACAAAGAAG	120
Db	68	AACATGGTTGCACCTTTTCACCTGGCTTTAAGTCAGCTGCTCATTCCCTGTTTCAAGGAAG	127
Qy	121	CAAAACCTTGACATCACCTTCATTTGCTAGCAATGGTGGAAAGAGTTAGCTGTCATGCAGGTTG	180
Db	128	CAAAACCTTGACATCACCTTCATTTGCCAGCAACGCGGAAGAGTCAATGTCATGCAGGTTG	187
Qy	181	TGGC	184
Db	188	TGGC	191

```

1  TITLE OF INVENTION: PROTEIN FUSION ENZYME
2
3  NUMBER OF SEQUENCES: 9
4
5  CORRESPONDENCE ADDRESS:
6
7  ADDRESSEE: Pennie & Edmonds
8  STREET: 2730 Sand Hill Road
9  CITY: Menlo Park
10 STATE: California
11
12 COUNTRY: U.S.A.
13
14 ZIP: 94025
15
16 COMPUTER READABLE FORM:
17
18 MEDIUM TYPE: Floppy disk
19
20 COMPUTER: IBM PC compatible
21
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23
24 SOFTWARE: Patent in Release #1.0,
25 SOFTWARE: Version #1.25
26

```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,483B
; FILING DATE: No. 552909ember 12, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7/857,602
; FILING DATE: March 30, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 923,692
; FILING DATE: July 31, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 600,244
; FILING DATE: October 22, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 641,617
; FILING DATE: January 16, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 737,899
; FILING DATE: July 26, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM:
; IMMEDIATE SOURCE:
; CLONE:
; FEATURE:
;
US-08-152-483B-8

Query Match 19.5%; Score 133.2; DB 2; Length 1442;
Best Local Similarity 84.3%; Pred. No. 5.9e-36;
Matches 150; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 1 ATGGCTTCCTGTGTCATTTCTTCAGCAGCTGTGCGCACACGACGCAATGTTACACAAGCT 60
Db 9 ATGGCTTCCTGTGTCATTTCTTCAGCAGCTGTGCGCACACGACGCAATGTTGCTCAAGCT 68

Qy 61 AGCATGGTTGCACCTTTCAGTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAGAAG 120
Db 69 AACATGGTTGCACCTTTCAGTGGCTTAAGTCAGTGGCTCATTCCTGTTTCAAGGAAG 128

Qy 121 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 178
Db 129 CAAAACCTTGACATCATTCCATTCCTAGCAACGCGGAAGAGTGCATGCAGCG 186

RESULT 11
US-08-095-726-12
; Sequence 12, Application US/08095726
; Patent No. 5530188
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L
; APPLICANT: Brinkhaus, Friedhelm L
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H
; APPLICANT: Yarger, James G
; APPLICANT: Yen, Huel-Che B
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in
```

```
; TITLE OF INVENTION: Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept
; STREET: 200 E Randolph St
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,726
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/785,566
; FILING DATE: 30-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5530188val B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 3128567180
; TELEFAX: 3128564972
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-095-726-12

Query Match 19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1 ATGGCTTCCTGTGTCATTTCTTCAGCAGCTGTGCGCACACGACGCAATGTTTACACAAGCT 60
Db 1 ATGGCTTCCTGTGTCATTTCTTCAGCAGCTGTGCGCACACGACGCAATGTTGCTCAAGCT 60

Qy 61 AGCATGGTTGCACCTTTCAGTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAGAAG 120
Db 61 AACATGGTTGCGCCTTTTCACCTGGCCTTAAGTCAGTGGCTCATTCCTGTTTCAAGGAAG 120

Qy 121 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177
Db 121 CAAAACCTTGACATCATTCCATTCCTAGCAACGCGGAAGAGTGCATGCAG 177

RESULT 12
US-08-096-043-12
; Sequence 12, Application US/08096043
; Patent No. 5530189
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L
; APPLICANT: Brinkhaus, Friedhelm L
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H
; APPLICANT: Yarger, James G
; APPLICANT: Yen, Huel-Che B
; TITLE OF INVENTION: Lycopene Biosynthesis in
; TITLE OF INVENTION: Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept
; STREET: 200 E Randolph St
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60680-0703
```

```
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.24
/ CURRENT APPLICATION DATA:
/ FILING DATE: 22-JUL-1993
/ APPLICATION NUMBER: US/08/096,043
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/785,568
/ FILING DATE: 30-OCT-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Galloway, No. 5530189val B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 3128567180
/ TELEFAX: 3128564972
/ INFORMATION FOR SEQ ID NO: 12:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 177 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-096-043-12

Query Match          19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCTCAATTTCTTCAGCAGCTGTTGCCACAGCGCAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCTCTCAAGTTCTTCTCTCGCAGCAGTTGCCACCGCGCAGCAATGTTCTCAAGCT 60

QY 61 AGCATGGTTGCACCTTTCACTGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db 61 AACATGGTGGCGCTTTCACTGGCCTTAAAGTCAGCTGCCTCAATTCCTGTTTCAAGGAAG 120

QY 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177
Db 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177

RESULT 13
US-08-093-577-8
/ Sequence 8, Application US/08093577
/ Patent No. 5545816
/ GENERAL INFORMATION:
/ APPLICANT: Ausich, Rodney L
/ APPLICANT: Brinkhaus, Friedhelm L
/ APPLICANT: Mukharji, Indrani
/ APPLICANT: Proffitt, John H
/ APPLICANT: Yarger, James G
/ APPLICANT: Yen, Huel-Che B
/ TITLE OF INVENTION: Phytoene Biosynthesis in
/ TITLE OF INVENTION: Genetically Engineered Hosts
/ NUMBER OF SEQUENCES: 43
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Amoco Corp., Patents and Licensing Dept
/ STREET: 200 E Randolph St
/ CITY: Chicago
/ STATE: IL
/ COUNTRY: USA
/ ZIP: 60680-0703
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/093,577
/ FILING DATE: 19-JUL-1993
/ CLASSIFICATION: 435
```

```
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/785,569
/ FILING DATE: 30-OCT-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Galloway, No. 5545816val B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 3128567180
/ TELEFAX: 3128564972
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 177 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-093-577-8

Query Match          19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCTCAATTTCTTCAGCAGCTGTTGCCACAGCGCAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCTCTCAAGTTCTTCTCTCGCAGCAGTTGCCACCGCGCAGCAATGTTCTCAAGCT 60

QY 61 AGCATGGTTGCACCTTTCACTGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db 61 AACATGGTGGCGCTTTCACTGGCCTTAAAGTCAGCTGCCTCAATTCCTGTTTCAAGGAAG 120

QY 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177
Db 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177

RESULT 14
US-08-331-004A-6
/ Sequence 6, Application US/08331004A
/ Patent No. 5618988
/ GENERAL INFORMATION:
/ APPLICANT: Hauptmann, Randal
/ APPLICANT: Eschenfeldt, William H
/ APPLICANT: English, Jami
/ APPLICANT: Brinkhaus, Friedhelm L
/ TITLE OF INVENTION: Enhanced Carotenoid Accumulation
/ TITLE OF INVENTION: in Storage Organs of Genetically
/ TITLE OF INVENTION: Engineered Plants
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Amoco Corporation, Law Dept
/ STREET: 55 Shuman Boulevard, Suite 600
/ CITY: Naperville
/ STATE: IL
/ COUNTRY: USA
/ ZIP: 60563-8437
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/331,004A
/ FILING DATE:
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Galloway, No. 5618988val B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 7087172447
/ TELEFAX: 7087172430
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 177 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-331-004A-6

Query Match      19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1 ATGGCTTCTCTGTCTTCATTTCTTCAGCAGTGTGGCCACACGAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCTCTAGTTCTTTCTCTCTCAGCAGTGTGCCACCGCAGCAATGTTGCTCAAGCT 60

Qy 61 AGCATGGTTGCACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db 61 AACATGGTGGCCCTTTCACTGGCCTTAAAGTCAGCTGCTCATTTCCCTGTTTCAAGGAAG 120

Qy 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTTAGCTGCATGCAG 177
Db 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTGCATGCATGCAG 177

RESULT 15
US-08-096-623A-12
; Sequence 12, Application US/08096623A
; Patent No. 5684238
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brinkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Hueli-Chen B.
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
; TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welsh & Katz, Ltd.
; STREET: 120 S. Riverside Plaza, 22nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,623A
; FILING DATE: 22-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,061
; FILING DATE: 09-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,921
; FILING DATE: 28-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/562,674
; FILING DATE: 03-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/525,551
; FILING DATE: 18-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/487,613
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; REFERENCE/DOCKET NUMBER: AMO-006.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 655-1500
; TELEFAX: (312) 655-1501
```

```
;
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; MAP POSITION: 1 to 177
; UNITS: bp
US-08-096-623A-12

Query Match      19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1 ATGGCTTCTCTGTCTTCATTTCTTCAGCAGCTGTGGCCACACGAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCTCTAGTTCTTTCTCTCTCAGTTCCTCTCTGACAGCTGTGCCACCGCAGCAATGTTGCTCAAGCT 60

Qy 61 AGCATGGTTGCACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db 61 AACATGGTGGCCCTTTCACTGGCCTTAAAGTCAGCTGCTCATTTCCCTGTTTCAAGGAAG 120

Qy 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTTAGCTGCATGCAG 177
Db 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTGCATGCATGCAG 177

Search completed: May 29, 2006, 12:19:38
Job time : 324 secs
```

GenCore version 5.1.8
 Copyright (c) 1993 - 2006 Bioceleration Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: May 30, 2006, 01:22:26 ; Search time 1263 Seconds
 (without alignments)
 6654.586 Million cell updates/sec

Title: US-10-718-311-7
 Perfect score: 684
 Sequence: 1 atggcttcctctgtcatttc.....cggcgctcacggtgtactaa 684

Scoring table: IDENTITY NUC
 Gapop 10.0 , Gapext 1.0

Searched: 18922170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq.*
- 15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq.*
- 16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	100.0	684	3	US-09-855-341-7
2	684	100.0	684	3	US-09-896-866B-16
3	684	100.0	684	7	US-10-359-369-41
4	684	100.0	684	8	US-10-699-050-16
5	684	100.0	684	8	US-10-718-311-7
6	684	100.0	684	9	US-10-462-162-33
7	501	73.2	3452	9	US-10-462-162-47
8	499.4	73.0	1971	9	US-10-462-162-54
9	498	72.8	498	9	US-10-462-162-29
10	495	72.4	495	3	US-09-855-341-3
11	495	72.4	495	3	US-09-896-866B-12
12	495	72.4	495	7	US-10-359-369-37
13	495	72.4	495	8	US-10-699-050-12
14	495	72.4	495	8	US-10-718-311-3
15	472.8	69.1	1207	10	US-10-450-763-26139
16	144	21.1	718	10	US-10-487-901-5765
17	144	21.1	736	10	US-10-487-901-1704

18	144	21.1	736	10	US-10-487-901-5043	Sequence 5043, Ap
19	142.4	20.8	297	10	US-10-487-901-7209	Sequence 7209, Ap
20	142.4	20.8	377	10	US-10-487-901-3504	Sequence 3504, Ap
21	142.4	20.8	489	10	US-10-487-901-3503	Sequence 3503, Ap
22	142.4	20.8	504	10	US-10-487-901-7207	Sequence 7207, Ap
23	140.8	20.6	204	3	US-09-839-477-3	Sequence 3, Appli
24	140.8	20.6	204	8	US-10-758-064-3	Sequence 3, Appli
25	137.6	20.1	668	10	US-10-487-901-7208	Sequence 7208, Ap
26	129.2	18.9	174	3	US-09-854-286-15	Sequence 15, Appli
27	128.8	18.8	169	6	US-10-165-420-3	Sequence 3, Appli
28	124.6	18.2	683	10	US-10-487-901-1713	Sequence 1713, Ap
29	124.6	18.2	737	10	US-10-487-901-5045	Sequence 5045, Ap
30	124.6	18.2	737	10	US-10-487-901-5766	Sequence 5766, Ap
31	121.4	17.7	738	10	US-10-487-901-1716	Sequence 1716, Ap
32	121.4	17.7	753	10	US-10-487-901-5764	Sequence 5764, Ap
33	121.4	17.7	754	10	US-10-487-901-5028	Sequence 5028, Ap
34	119.8	17.5	714	10	US-10-487-901-1703	Sequence 1703, Ap
35	119.8	17.5	736	10	US-10-487-901-5044	Sequence 5044, Ap
36	115	16.8	704	9	US-10-425-115-85509	Sequence 85509, A
37	113	16.5	151	7	US-10-321-434-4	Sequence 4, Appli
38	111.8	16.3	654	10	US-10-487-901-5042	Sequence 5042, Ap
39	111	16.2	180	10	US-10-487-901-88	Sequence 88, Appli
40	97.2	14.2	149	3	US-09-839-477-5	Sequence 5, Appli
41	97.2	14.2	149	8	US-10-758-064-5	Sequence 5, Appli
42	85.2	12.5	363	3	US-09-770-791-468	Sequence 468, App
43	85.2	12.5	711	3	US-09-910-664-32	Sequence 32, Appli
44	85.2	12.5	711	8	US-10-333-184-31	Sequence 31, Appli
45	85.2	12.5	739	8	US-10-333-184-182	Sequence 182, App

ALIGNMENTS

RESULT 1

US-09-855-341-7
 ; Sequence 7, Application US/09855341
 ; Patent No. US20020002715A1
 ; GENERAL INFORMATION:
 ; APPLICANT: VIITANEN, PAUL V.
 ; APPLICANT: MEYER, KNUZ
 ; APPLICANT: VAN DYK, DREW
 ; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
 ; TITLE OF INVENTION: IN GREEN PLANTS
 ; FILE REFERENCE: BC1015 US NA
 ; CURRENT APPLICATION NUMBER: US/09/855,341
 ; CURRENT FILING DATE: 2001-05-15
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: MICROSOFT OFFICE 97
 ; SEQ ID NO 7
 ; LENGTH: 684
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
 US-09-855-341-7

Query Match	100.0%	Score 684;	DB 3;	Length 684;
Best Local Similarity	100.0%;	Pred. No. 1.9e-222;	Indels 0;	Gaps 0;
Matches 684;	Conservative 0;	Mismatches 0;		
QY	1	ATGGCTTCTCTGTGTCATTTCTTTCAGCAGCTGTTGGCCACACGAGCAATGTTACACAAGCT	60	
Db	1	ATGGCTTCTCTGTGTCATTTCTTTCAGCAGCTGTTGGCCACACGAGCAATGTTACACAAGCT	60	
QY	61	AGCATGGTTGCACCTTTCATCGTCTCAAAATCTTTCAGCCACCTTTCCTCTTCAAAAGAAG	120	
Db	61	AGCATGGTTGCACCTTTCATCGTCTCAAAATCTTTCAGCCACCTTTCCTCTTCAAAAGAAG	120	
QY	121	CAAAACCTTGACATCATTCCATTCTAGCAATGTTGGAGAGTTAGCTGCATGCAGGTTG	180	
Db	121	CAAAACCTTGACATCATTCCATTCTAGCAATGTTGGAGAGTTAGCTGCATGCAGGTTG	180	
QY	181	TGGCATATGTACACACCGCGGTTAACGCAACTCGCTGCGCTGCTATTGTTAAAGAGATC	240	

Db 1 ATGGCTTCCTCTGTCAATTTCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAGCT 60
QY 61 AGCATGGTTGCACCTTTCTACTGTGCTCAATCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAG 120
Db 61 AGCATGGTTGCACCTTTCTACTGTGCTCAATCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAG 120
QY 121 CAAACCTTTGACATCACTTCTCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGGTG 180
Db 121 CAAACCTTTGACATCACTTCTCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGGTG 180
QY 181 TGGCATATGTCAACCCCGGTTAAACGCACTGCTGCTGGAGATTTCCATGACAAACAGT 240
Db 181 TGGCATATGTCAACCCCGGTTAAACGCACTGCTGCTGGAGATTTCCATGACAAACAGT 240
QY 241 CTGCGCTCGATCCGCAACTGCTGCACTGCTGCTGGAGATTTCCATGACAAACAGT 300
Db 241 CTGCGCTCGATCCGCAACTGCTGCACTGCTGCTGGAGATTTCCATGACAAACAGT 300
QY 301 TTTGAAACAGCAGGAGAAAACGGTAAAGCGTGAACGATGATCCGCAAGGGTTTGTGCGAGCAG 360
Db 301 TTTGAAACAGCAGGAGAAAACGGTAAAGCGTGAACGATGATCCGCAAGGGTTTGTGCGAGCAG 360
QY 361 AATGAAATCCCGAAGAACTGCGGCTGCTGCGGAAAGAGTCTGTTACTGTTTACCGTGAA 420
Db 361 AATGAAATCCCGAAGAACTGCGGCTGCTGCGGAAAGAGTCTGTTACTGTTTACCGTGAA 420
QY 421 ATTTTGTATGTGCGCATGTTGTAACCGTGGTAAACCGTGGTAAACCGTGGTAAACCGTGGTAA 480
Db 421 ATTTTGTATGTGCGCATGTTGTAACCGTGGTAAACCGTGGTAAACCGTGGTAAACCGTGGTAA 480
QY 481 ACGTTAAGCGGCGGAGCTGCGGTTACAAAATTCGGTAAACCGGTTAGGACGCTAT 540
Db 481 ACGTTAAGCGGCGGAGCTGCGGTTACAAAATTCGGTAAACCGGTTAGGACGCTAT 540
QY 541 CTGTTCAATCATCGACATTAACCGGAGCTTATTTAGATAGGCGGTGATGCGCGGCTG 600
Db 541 CTGTTCAATCATCGACATTAACCGGAGCTTATTTAGATAGGCGGTGATGCGCGGCTG 600
QY 601 TGGGGCGAGCTTCCCGCTGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTT 560
Db 601 TGGGGCGAGCTTCCCGCTGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTT 560
QY 661 TTACCGGCGTCACCGTTGTACTAA 684
Db 661 TTACCGGCGTCACCGTTGTACTAA 684

RESULT 4
US-10-699-050-16
; Sequence 16, Application US/10699050
; Publication NO. US20040142437A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/10/699,050
; CURRENT FILING DATE: 2003-10-30
; PRIOR APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature

; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protei
US-10-699-050-16
Query Match 100.0%; Score 684; DB 8; Length 684;
Best Local Similarity 100.0%; Pred. No. 1.9e-222;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTCCTCTGTCAATTTCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCCTCTGTCAATTTCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAGCT 60
QY 61 AGCATGGTTGCACCTTTCTACTGTGCTCAATCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAG 120
Db 61 AGCATGGTTGCACCTTTCTACTGTGCTCAATCTTCAGCAGCTGTTGCCACACGCAATGTTACACAAG 120
QY 121 CAAACCTTTGACATCACTTCTCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGGTG 180
Db 121 CAAACCTTTGACATCACTTCTCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGGTG 180
QY 181 TGGCATATGTCAACCCCGGTTAAACGCACTGCTGCTGGAGATTTCCATGACAAACAGT 240
Db 181 TGGCATATGTCAACCCCGGTTAAACGCACTGCTGCTGGAGATTTCCATGACAAACAGT 240
QY 241 CTGCGCTCGATCCGCAACTGCTGCACTGCTGCTGGAGATTTCCATGACAAACAGT 300
Db 241 CTGCGCTCGATCCGCAACTGCTGCACTGCTGCTGGAGATTTCCATGACAAACAGT 300
QY 301 TTTGAAACAGCAGGAGAAAACGGTAAAGCGTGAACGATGATCCGCAAGGGTTTGTGCGAGCAG 360
Db 301 TTTGAAACAGCAGGAGAAAACGGTAAAGCGTGAACGATGATCCGCAAGGGTTTGTGCGAGCAG 360
QY 361 AATGAAATCCCGAAGAACTGCGGCTGCTGCGGAAAGAGTCTGTTACTGTTTACCGTGAA 420
Db 361 AATGAAATCCCGAAGAACTGCGGCTGCTGCGGAAAGAGTCTGTTACTGTTTACCGTGAA 420
QY 421 ATTTTGTATGTGCGCATGTTGTAACCGTGGTAAACCGTGGTAAACCGTGGTAAACCGTGGTAA 480
Db 421 ATTTTGTATGTGCGCATGTTGTAACCGTGGTAAACCGTGGTAAACCGTGGTAAACCGTGGTAA 480
QY 481 ACGTTAAGCGGCGGAGCTGCGGTTACAAAATTCGGTAAACCGGTTAGGACGCTAT 540
Db 481 ACGTTAAGCGGCGGAGCTGCGGTTACAAAATTCGGTAAACCGGTTAGGACGCTAT 540
QY 541 CTGTTCAATCATCGACATTAACCGGAGCTTATTTAGATAGGCGGTGATGCGCGGCTG 600
Db 541 CTGTTCAATCATCGACATTAACCGGAGCTTATTTAGATAGGCGGTGATGCGCGGCTG 600
QY 601 TGGGGCGAGCTTCCCGCTGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTT 660
Db 601 TGGGGCGAGCTTCCCGCTGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTT 660
QY 661 TTACCGGCGTCACCGTTGTACTAA 684
Db 661 TTACCGGCGTCACCGTTGTACTAA 684

RESULT 5
US-10-718-311-7
; Sequence 7, Application US/10718311
; Publication NO. US20040143867A1
; GENERAL INFORMATION:
; APPLICANT: Viitanen, Paul V.
; APPLICANT: Meyer, Knut
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/10/718,311
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: US/09/855,341
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97

; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 47
; LENGTH: 3452
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of nucleic acid fragment encoding CPL (from
; OTHER INFORMATION: Escherichia coli), pHB 1-hydroxylase (from Candida
; OTHER INFORMATION: parapsilosis), and UGT72B1 (from Arabidopsis thaliana).
US-10-462-162-47

Query Match 73.2%; Score 501; DB 9; Length 3452;
Best Local Similarity 100.0%; Pred. No. 1.8e-159;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 184 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCTGAGGATCCATGACAAAACGTTTT 303
Db 1 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCTGAGGATCCATGACAAAACGTTTT 60
QY 244 GCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATCCATGACAAAACGTTTT 303
Db 61 GCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATCCATGACAAAACGTTTT 120
QY 304 GAACAGCAGGAGAAAACGGTACGCTGACGATCCGCGAAGGGTTTGTGAGCAGAAAT 363
Db 121 GAACAGCAGGAGAAAACGGTACGCTGACGATCCGCGAAGGGTTTGTGAGCAGAAAT 180
QY 364 GAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAAT 423
Db 181 GAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAAT 240
QY 424 TTGTTATGCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGTCACG 483
Db 241 TTGTTATGCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGTCACG 300
QY 484 TTAAGCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTG 543
Db 301 TTAAGCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTG 360
QY 544 TTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGCGTATGCCGGCTGTGG 603
Db 361 TTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGCGTATGCCGGCTGTGG 420
QY 604 GGGCGACGTTCCCGCTCGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTA 663
Db 421 GGGCGACGTTCCCGCTCGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTA 480
QY 664 CCGGCGTCACCGTTGTACTAA 684
Db 481 CCGGCGTCACCGTTGTACTAA 501

RESULT 8
US-10-462-162-54
; Sequence 54, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 1971
; TYPE: DNA

; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleic acid sequence of a nucleic acid fragment inserted into an
; OTHER INFORMATION: expression vector pET29a encoding CPL (from Escherichia coli) an
; OTHER INFORMATION: pHB 1-hydroxylase (from Candida parapsilosis).
US-10-462-162-54

Query Match 73.0%; Score 499.4; DB 9; Length 1971;
Best Local Similarity 99.8%; Pred. No. 4.6e-159;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 184 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCTGAGGATCCATGACAAAACGTTTT 243
Db 1 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCTGAGGATCCATGACAAAACGTTTT 60
QY 244 GCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATCCATGACAAAACGTTTT 303
Db 61 GCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATCCATGACAAAACGTTTT 120
QY 304 GAACAGCAGGAGAAAACGGTACGCTGACGATCCGCGAAGGGTTTGTGAGCAGAAAT 363
Db 121 GAACAGCAGGAGAAAACGGTACGCTGACGATCCGCGAAGGGTTTGTGAGCAGAAAT 180
QY 364 GAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAAT 423
Db 181 GAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAAT 240
QY 424 TTGTTATGCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGTCACG 483
Db 241 TTGTTATGCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGTCACG 300
QY 484 TTAAGCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTG 543
Db 301 TTAAGCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTG 360
QY 544 TTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGCGTATGCCGGCTGTGG 603
Db 361 TTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGCGTATGCCGGCTGTGG 420
QY 604 GGGCGACGTTCCCGCTCGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTA 663
Db 421 GGGCGACGTTCCCGCTCGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTA 480
QY 664 CCGGCGTCACCGTTGTACTAA 684
Db 481 CCGGCGTCACCGTTGTACTAA 501

RESULT 9
US-10-462-162-29
; Sequence 29, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-29

Query Match 72.8%; Score 498; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 6.9e-159;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; TYPE: DNA

```
QY 187 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
Db 1 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATCACAACAGTTTTCGA 306
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATCACAACAGTTTTCGA 120
QY 307 CAGCAGGGAACCGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGA 366
Db 121 CAGCAGGGAACCGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGA 180
QY 367 ATCCCGGAAGACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 426
Db 181 ATCCCGGAAGACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 240
QY 427 TTATGTGCGGATGGTGAACCGTGGCTTACCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 486
Db 241 TTATGTGCGGATGGTGAACCGTGGCTTACCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 300
QY 487 AGCGGCGGAGCTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 486
Db 241 TTATGTGCGGATGGTGAACCGTGGCTTACCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 300
QY 487 AGCGGCGGAGCTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 546
Db 301 AGCGGCGGAGCTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 360
QY 547 ACATCATCGACATTAACCGGAGCTTTATTTAGATAGGCGGTGATGCGGCGCTGTGGGG 606
Db 361 ACATCATCGACATTAACCGGAGCTTTATTTAGATAGGCGGTGATGCGGCGCTGTGGGG 420
QY 607 CGAGTTCGCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 666
Db 421 CGAGTTCGCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480
QY 667 GCGTCACCGTTGTACTAA 684
Db 481 GCGTCACCGTTGTACTAA 498
```

```
RESULT 10
US-09-855-341-3
; Sequence 3, Application US/09855341
; Patent No. US2002002715A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUD
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: BCI015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-09-855-341-3
```

```
Query Match 72.4%; Score 495; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 7.3e-158;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
Db 1 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATCACAACAGTTTTCGA 306
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATCACAACAGTTTTCGA 120
QY 307 CAGCAGGGAACCGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGA 366
```

```
Db 121 CAGCAGGGAACCGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGA 180
QY 367 ATCCCGGAAGACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 426
Db 181 ATCCCGGAAGACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 240
QY 427 TTATGTGCGGATGGTGAACCGTGGCTTACCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 486
Db 241 TTATGTGCGGATGGTGAACCGTGGCTTACCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 300
QY 487 AGCGGCGGAGCTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 546
Db 301 AGCGGCGGAGCTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 360
QY 547 ACATCATCGACATTAACCGGAGCTTTATTTAGATAGGCGGTGATGCGGCGCTGTGGGG 606
Db 361 ACATCATCGACATTAACCGGAGCTTTATTTAGATAGGCGGTGATGCGGCGCTGTGGGG 420
QY 607 CGAGTTCGCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 666
Db 421 CGAGTTCGCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480
QY 667 GCGTCACCGTTGTACT 681
Db 481 GCGTCACCGTTGTACT 495
RESULT 11
US-09-896-866B-12
; Sequence 12, Application US/09896866B
; Patent No. US20020151002A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylgucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12
Query Match 72.4%; Score 495; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 7.3e-158;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
Db 1 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATCACAACAGTTTTCGA 306
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATCACAACAGTTTTCGA 120
QY 307 CAGCAGGGAACCGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGA 366
Db 121 CAGCAGGGAACCGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGA 180
QY 367 ATCCCGGAAGACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 426
Db 181 ATCCCGGAAGACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 240
QY 427 TTATGTGCGGATGGTGAACCGTGGCTTACCGGTGCTACCGTGGTTCCTGTGTCAACGTTA 486
```


QY 667 GCGTCACCGTTGTAC 681
Db 481 GCGTCACCGTTGTAC 495

RESULT 14

US-10-718-311-3
; Sequence 3, Application US/10718311
; Publication No. US20040143867A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: IN GREEN PLANTS
; CURRENT APPLICATION NUMBER: US/10/718,311
; PRIOR FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: US/09/855,341
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-10-718-311-3

Query Match 72.4%; Score 495; DB 8; Length 495;
Best Local Similarity 100.0%; Pred. No. 7,38-158;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGAGATCCCTGCC 246
Db 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 60

QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 306
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 120

QY 307 CAGCAGGGGAAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGAGCAATGAA 366
Db 121 CAGCAGGGGAAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGAGCAATGAA 180

QY 367 ATCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426
Db 181 ATCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240

QY 427 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTGACCGTCTGTTCTGTGTCAACGTTA 486
Db 241 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTGACCGTCTGTTCTGTGTCAACGTTA 300

QY 487 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACGCCGTTAGGACGCTATCTGTTTC 546
Db 301 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACGCCGTTAGGACGCTATCTGTTTC 360

QY 547 ACATCATCGACATTAACCCGGACTTTTATGAGATAGGCGGCTGATGCCGGGCTGTGGGG 606
Db 361 ACATCATCGACATTAACCCGGACTTTTATGAGATAGGCGGCTGATGCCGGGCTGTGGGG 420

QY 607 CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666
Db 421 CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480

QY 667 GCGTCACCGTTGTAC 681
Db 481 GCGTCACCGTTGTAC 495

RESULT 15

US-10-450-763-26139
; Sequence 26139, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 26139
; LENGTH: 1207
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (893)..(925)
; OTHER INFORMATION: 81% homologous to Escherichia coli 4-hydroxybenzoate
; OTHER INFORMATION: synthetase, accession number X57434, Smith-Waterman Score=49.
US-10-450-763-26139

Query Match 69.1%; Score 472.8; DB 10; Length 1207;
Best Local Similarity 99.2%; Pred. No. 4,5e-150;
Matches 496; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246
Db 652 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 711

QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 306
Db 712 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 771

QY 307 CAGCAGGGGAAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGAGCAATGAA 366
Db 772 CAGCAGGGGAAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGCGAGCAATGAA 831

QY 367 ATCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426
Db 832 ATCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 891

QY 427 T-TATGTCCGATGGTGAACCGTGGCTTGGCC- GGTCGTACCGTCTGTTCTGTGTCAACGT 484
Db 892 TGTATGTCCGATGGCGNAACCGCGTTGCCGGTCTGTAACCGTCTGTGTGTCAACGT 951

QY 485 TAAGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACGCCGTTAGGACGCTATCTGT 544
Db 952 TAAGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACGCCGTTAGGACGCTATCTGT 1011

QY 545 TCACATCATCGACATTAACCCGGACTTTATGAGATAGGCGGCTGATGCCGGGCTGTGGG 604
Db 1012 TCACATCATCGACATTAACCCGGACTTTATGAGATAGGCGGCTGATGCCGGGCTGTGGG 1071

QY 605 GGCGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGTCGTAACAGAACTGTTTTTAC 664
Db 1072 GGCGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGTCGTAACAGAACTGTTTTTAC 1131

QY 665 CGCGCTCACCGTTGTACTAA 684
Db 1132 CGCGCTCACCGTTGTACTAA 1151

Search completed: May 30, 2006, 01:43:38
Job time : 1265 secs

November 2005

Published_Applications_Nucleic Acid and Published_Applications_Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).

Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2006, 01:24:28 ; Search time 44 Seconds
(without alignments)
1754.839 Million cell updates/sec

Title: US-10-718-311-7
Perfect score: 684
Sequence: 1 atggcttcctctgtcatttc.....cggcgtaaccgtgtactaa 684

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 242596 seqs, 56442199 residues

Total number of hits satisfying chosen parameters: 485192

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA_New.*

- 1: /EMC Celerra_SIDS3/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 2: /EMC Celerra_SIDS3/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 3: /EMC Celerra_SIDS3/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 4: /EMC Celerra_SIDS3/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 5: /EMC Celerra_SIDS3/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 6: /EMC Celerra_SIDS3/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 7: /EMC Celerra_SIDS3/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 8: /EMC Celerra_SIDS3/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.2	12.5	718	6	US-10-953-349-10521
2	82	12.0	743	6	US-10-953-349-8894
3	74.4	10.9	799	6	US-10-953-349-9569
4	35.6	5.2	792	7	US-11-217-529-6134
5	32	4.7	2144	7	US-11-293-697-312
6	31.2	4.6	336	7	US-11-217-529-77187
7	30.8	4.5	2586	7	US-11-217-529-1832
8	30.4	4.4	807	6	US-10-953-349-40123
9	30	4.4	1801	6	US-10-953-349-7256
10	29.8	4.4	711	7	US-11-217-529-76273
11	29.8	4.4	3449	7	US-11-293-697-830
12	29.2	4.3	459	7	US-11-301-554-1043
13	29	4.2	1584	7	US-11-217-529-80201
14	28.8	4.2	450	7	US-11-301-554-1056
15	28.8	4.2	9399	7	US-11-236-836-2
16	28.8	4.2	11918	7	US-11-236-836-22
17	28.8	4.2	12023	7	US-11-236-836-20
18	28.8	4.2	13042	7	US-11-236-836-17
19	28.8	4.2	13042	7	US-11-236-836-23
20	28.8	4.2	13147	7	US-11-236-836-18
21	28.8	4.2	13147	7	US-11-236-836-25
22	28.8	4.2	901	6	US-10-953-349-1271
23	28.6	4.2	2160	7	US-11-293-697-438
24	28.4	4.2	855	7	US-11-217-529-77669
25	28.4	4.2	1107	7	US-11-217-529-78001

C	26	28.2	4.1	518	6	US-10-488-619-2571	Sequence 2571, Ap
C	27	28.2	4.1	1451	6	US-10-953-349-10879	Sequence 10879, A
C	28	28	4.1	2271	6	US-10-505-928-792	Sequence 792, App
C	29	28	4.1	3240	7	US-11-217-529-1227	Sequence 1227, Ap
C	30	27.6	4.0	1088	6	US-10-953-349-9573	Sequence 9573, Ap
C	31	27.6	4.0	3738	7	US-11-217-529-473	Sequence 473, Appl
C	32	27.6	4.0	3953	6	US-10-473-173-56	Sequence 56, Appl
C	33	27.4	4.0	1435	6	US-10-953-349-13607	Sequence 13607, A
C	34	27.4	4.0	1714	6	US-10-953-349-32909	Sequence 32909, A
C	35	27.2	4.0	726	6	US-10-953-349-40081	Sequence 40081, A
C	36	27.2	4.0	1320	7	US-11-311-555-5	Sequence 5, Appli
C	37	27.2	4.0	1320	7	US-11-311-561-5	Sequence 5, Appli
C	38	27.2	4.0	1320	7	US-11-101-316-155	Sequence 155, App
C	39	27.2	4.0	2568	7	US-11-217-529-75586	Sequence 75586, A
C	40	27	3.9	612	7	US-11-217-529-174132	Sequence 174132, A
C	41	27	3.9	1157	6	US-10-953-349-35133	Sequence 35133, A
C	42	27	3.9	1760	6	US-10-953-349-11108	Sequence 11108, A
C	43	27	3.9	138941	6	US-10-489-730-10	GENERAL INFORMATI
C	44	26.8	3.9	549	6	US-10-953-349-11985	Sequence 11985, A
C	45	26.8	3.9	774	7	US-11-217-529-174376	Sequence 174376,

ALIGNMENTS

RESULT 1

US-10-953-349-10521
; Sequence 10521, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10521
; LENGTH: 718
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-10521

Query Match 12.5%; Score 85.2; DB 6; Length 718;
Best Local Similarity 70.7%; Pred. No. 1e-19;
Matches 130; Conservative 0; Mismatches 48; Indels 6; Gaps 1;

QY	1	ATGGCTTCCTCTGTCTATTTCTTCAGCAGCTGTTCACACGACGCAATGTTACACAGCT	60
Db	32	ATGGCTTCCTCTATGCTCTCTCTGCGGTGTGTAC-----CTCCCCGGCTCAAGCC	85
QY	61	AGATGGTTGCACTTTCATCTGCTCTCAAAATCTTACGCCACTTTCCCTGTTACAAAGAAG	120
Db	86	ACCATGGTCGCTCACTACTGTTTGAAGTCACTCGCTCTTTCCCGGTCACCCGCAAG	145
QY	121	CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGAAGAGTTCATCGTCAGTGCTG	180
Db	146	GCCAAACAGACATTAATCTTCCATCACANGAATGGGGGAAGATTAGTTCATGAAGGTG	205
QY	181	TGCG 184	
Db	206	TGCG 209	

RESULT 2

US-10-953-349-8894
; Sequence 8894, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY

```
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8894
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-8894

Query Match          12.0%; Score 82; DB 6; Length 743;
Best Local Similarity 69.6%; Pred. No. 1.3e-18;
Matches 128; Conservative 0; Mismatches 50; Indels 6; Gaps 1;

; Qy 1 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTCACACGACGCAATGTTACACAAGCT 60
; Db ATGGCTTCTCTATGCTCTCTCCGCGCTGTGGTTACATCC-----CCGGCTCAGGCC 106
; Qy 61 AGCATGGTTGACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
; Db ATGGCTTCTCTATGCTCTCTCCGCGCTGTGGTTACATCC-----CCGGCTCAGGCC 166
; Qy 107 ACCATGGTGGCTCCATTACCGGCTTGAAGTCATCGCTGCATTCCTCCGCTCACCCGCAAG 166
; Db ATGGCTTCTCTATGCTCTCTCCGCGCTGTGGTTACATCC-----CCGGCTCAGGCC 180
; Qy 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGTTGAAGAGTTAGCTGCATGCAGGTG 180
; Db ACCAACAAGGACATCACTTCCATTCGACGACGCGGGAAGAGTTAGCTGCATGAAGGTG 226
; Qy 181 TGGC 184
; Db TGGC 230
; Qy 227 TGGC 230

RESULT 3
US-10-953-349-9569
; Sequence 9569, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9569
; LENGTH: 799
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-9569

Query Match          10.9%; Score 74.4; DB 6; Length 799;
Best Local Similarity 66.7%; Pred. No. 5.6e-16;
Matches 124; Conservative 0; Mismatches 56; Indels 6; Gaps 1;

; Qy 1 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTCACACGACGCAATGTTACACAAGCT 60
; Db ATGGCTTCTCTATGCTCTCTCCGCTACTATGTTGC-----CTCTCCGGCTCAGGCC 100
; Qy 61 AGCATGGTTGACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
; Db ACTATGGTGGCTTCCCTTTCACGAGCTTAAGTCTCGCTGCTTCCAGCCACCCGCAAG 160
; Qy 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGTTGAAGAGTTAGCTGCATGCAGGTG 180
; Db GCTAACACGACATTACTTCCATCAAGCAACGCGGAAGAGTTAACTGTCATGCAGGTG 220
; Qy 181 TGGCAT 186
; Db TGGCCT 226

RESULT 4
US-10-953-349-9569
; Sequence 9569, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9569
; LENGTH: 799
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-9569

Query Match          10.9%; Score 74.4; DB 6; Length 799;
Best Local Similarity 66.7%; Pred. No. 5.6e-16;
Matches 124; Conservative 0; Mismatches 56; Indels 6; Gaps 1;

; Qy 1 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTCACACGACGCAATGTTACACAAGCT 60
; Db ATGGCTTCTCTATGCTCTCTCCGCTACTATGTTGC-----CTCTCCGGCTCAGGCC 100
; Qy 61 AGCATGGTTGACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
; Db ACTATGGTGGCTTCCCTTTCACGAGCTTAAGTCTCGCTGCTTCCAGCCACCCGCAAG 160
; Qy 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGTTGAAGAGTTAGCTGCATGCAGGTG 180
; Db GCTAACACGACATTACTTCCATCAAGCAACGCGGAAGAGTTAACTGTCATGCAGGTG 220
; Qy 181 TGGCAT 186
; Db TGGCCT 226

RESULT 5
US-11-293-697-312
; Sequence 312, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 312
; LENGTH: 2144
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-312

Query Match          4.7%; Score 32; DB 7; Length 2144;
Best Local Similarity 62.5%; Pred. No. 0.35;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

; Qy 2 TGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTCACACGACGCAATGTTACACAAGCTA 61
; Db TGGTTCATCTGCTTTTGATTTCAGAGCTCTTTCCTACCCAGCAAGCTACACTTCTT 1736
; Qy 62 GCATGGTTGCACCTTTCACT 81
; Db GCCTTCTTCTATTTTCT 1756
```

```

, , TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
, , FILE REFERENCE: S-38-285
, , CURRENT APPLICATION NUMBER: US/11/217,529
, , CURRENT FILING DATE: 2005-09-02
, , PRIOR APPLICATION NUMBER: US 10/932,182
, , PRIOR FILING DATE: 2004-09-02
, , NUMBER OF SEQ ID NOS: 197023
, , SOFTWARE: PatentIn version 3.3
, , SEQ ID NO 77187
, , LENGTH: 336
, , TYPE: DNA
, , ORGANISM: Saccharomyces pastorianus
US-11-217-529-77187

```

```

; SEQ ID NO 40123
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-40123

```

Qy	219	GCTGGCGCTATTGTAAA	234
Db	156	GCCATGTATCTGAAA	141

Query Match 4.4%; Score 30; DB 6; Length 1801;
Best Local Similarity 48.8%; Pred. No. 1.6;
Matches 81; Conservative 0; Mismatches 85; Indels 0; Gaps 0

Qy 6 TTCTCTGTCTATTTCTTCAGCAGCTGTGGCACGCGCAATGTTACACAAGCTAGCAT 65

Db 20 TTCCTTGTTCAACCTCTGAGAGATTACGGATAGTGAAGAGAGCTACAAAATTTCCCAA 79
QY 66 GGTTCACCTTTCTACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAGCAAAA 125
Db 80 TTTATCAAAAAATAGAAATTTCTTACTTTTCTTCTTTTCAAAATTCATTTCTCCCAAAAAAATA 139
QY 126 CTTGACATCACTTCATCTGTAGCAATGGTGAAGAGATTAGCTGC 171
Db 140 TCTTGTGAAACTGCAACTGTGATCCATGGCGGAGGATATGGCTGC 185

RESULT 10
US-11-217-529-76273
; Sequence 76273, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 76273
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-76273

Query Match 4.4%; Score 29.8; DB 7; Length 711;
Best Local Similarity 49.1%; Pred. No. 1.1;
Matches 79; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 5 CTTCTCTGTCTATTCTTCAGCAGCTGTTGCCACAGCGCAATGTTACAAAGCTAGCA 64
Db 488 CTTCTGAGCTAAGTCTTCTCTGCTGCTCCAGCTCCACTGAAGCTAAGCACTTCTG 547
QY 65 TGGTTGACCTTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTTACAAAGAAGCAAA 124
Db 548 CTGCTCCAAGCTCCACTGCTGGTCCAAAGACTCTGCCATCTCTCAAAATTACCGATGGTCAA 607
QY 125 ACCTTGACATCACTTCCATGCTAGCAATGGTGGAGAGTT 165
Db 608 TCCAAAGCTACCAAGGCTGTTTCTGAGCAAACTGAAAACGGT 648

RESULT 11
US-11-293-697-830/c
; Sequence 830, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 830
; LENGTH: 3449
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-830

Query Match 4.4%; Score 29.8; DB 7; Length 3449;
Best Local Similarity 54.0%; Pred. No. 2.7;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 80 CTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTTACAAAGAAGCAAAACCTTTGACATCACTT 139
Db 1379 CTTCTTTAACGTTATCACACATACACATACAAAGAGGAGCAGCAGTACATCTACT 1320
QY 140 CCATTGTAGCAATGGTGGAAAGAGTTAGCTGCATGCAGGTGTGGCATATGTCA 192
Db 1319 CAAGTCAATGAAAAATAAAAGTAAAGATGCCATCAAAACTTGAATAAGTCA 1267

RESULT 12
US-11-301-554-1043
; Sequence 1043, Application US/11301554
; Publication No. US20060088527A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Durnam, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedgwick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McHabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C21
; CURRENT APPLICATION NUMBER: US/11/301,554
; CURRENT FILING DATE: 2005-12-13
; PRIOR APPLICATION NUMBER: US 10/283,017
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US 10/113,872
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 10/017,754
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 09/902,941
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 09/849,626
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 09/736,457
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/702,705
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: US 09/677,419
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/671,325
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/658,824
; PRIOR FILING DATE: 2000-09-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1043
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-301-554-1043

Query Match 4.3%; Score 29.2; DB 7; Length 459;
Best Local Similarity 50.7%; Pred. No. 1.3;
Matches 70; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 253 CCGCAACTGCTCGACTGGCTTCTGAGAGATTCCATGACAAACGTTTGAACAGCAG 312
Db 96 CAGGACCTGCTGGTTTCCCTGCTGCTGACAGAAATGGTGAACCTGGTGTAGGGAG 155

QY 313 GGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTGTTCGAGCAGAGTAATAATCCCC 372
Db 156 AAAGAGGGGCTCCGGGTGAGAAAGGTGAAGAGGCCCTCTCGGATTCGAGGACCCCTG 215
QY 373 GAAGAACTGCGCTGCTG 390
Db 216 GAGGTCTGACACTGCTG 233

RESULT 13

US-11-217-529-80201
; Sequence 80201, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 80201
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-80201

Query Match 4.2%; Score 29; DB 7; Length 1584;
Best Local Similarity 51.1%; Pred. No. 3.2;
Matches 68; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 289 ATGACAAAAAGCTTTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGG 348
Db 1 ATGCGGAGATTATGACACTAGCCATTAAAGCCACTAATACGGTGTGATGACATGGG 60
QY 349 TTGTGTCGACAGAAATAATCCCGAAGAACTGCCGCTGCTGCCGCGAAGAGTCTCGTTAC 408
Db 61 TTTGAAGAGCAATACCAAGTGACCAACTTTTACATCACTCCGCAATGATTTCATATTC 120
QY 409 TGGTTACGTGAAA 421
Db 121 TACTTTGATGATA 133

RESULT 14

US-11-301-554-1056/c
; Sequence 1056, Application US/11301554
; Publication No. US20060088527A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C21
; CURRENT APPLICATION NUMBER: US/11/301,554

; CURRENT FILING DATE: 2005-12-13
; PRIOR APPLICATION NUMBER: US 10/283,017
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US 10/113,872
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 10/017,754
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 09/902,941
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 09/849,626
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 09/736,457
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/702,705
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: US 09/677,419
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/671,325
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/658,824
; PRIOR FILING DATE: 2000-09-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1056
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 21, 22, 230, 232, 377, 391
; OTHER INFORMATION: n = A,T,C or G
US-11-301-554-1056

Query Match 4.2%; Score 28.8; DB 7; Length 450;
Best Local Similarity 50.0%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 253 CCGCAACTCTCTCGACTGGCTGTTCGTGGAGGATTCATGACAAAACTTTTGAACAGCAG 312
Db 364 CAGGACCTCTGTTTCCCTGGTCTCTCGACAGAAATGTTGAACTGGTGGTAAAGGAG 305
QY 313 GGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTGTTCGAGCAGAGTAATAATCCCC 372
Db 304 AAAGAGGGGCTCCGGTGAGAAAGGTGAAGGAGGCCCTCTCTGAGTTGCAGGACCCCTG 245
QY 373 GAAGAACTGCGCTGCTG 390
Db 244 GAGTTCTGGACNTCTG 227

RESULT 15

US-11-236-836-2/c
; Sequence 2, Application US/11236836
; Publication No. US20060105365A1
; GENERAL INFORMATION:
; APPLICANT: MARTIN, Annette
; APPLICANT: GHIBAUDO, David
; APPLICANT: COHEN, Lisette
; APPLICANT: LEMON, STANLEY M.
; TITLE OF INVENTION: Chimeric GB Virus B (GBV-B)
; FILE REFERENCE: UTSG:270US
; CURRENT APPLICATION NUMBER: US/11/236,836
; CURRENT FILING DATE: 2005-09-27
; PRIOR APPLICATION NUMBER: 60/613,266
; PRIOR FILING DATE: 2004-09-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 9399
; TYPE: DNA
; ORGANISM: GBV-A-like virus
US-11-236-836-2

Query Match 4.2%; Score 28.8; DB 7; Length 9399;
Best Local Similarity 69.6%; Pred. No. 10;
Matches 39; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 113 CAAGAAGCAAAACCTTGACATCACTTCCTAGCTAGCAATGGTGGAAAGATTAGC 168
|||||
Db 2720 CAAGAAGCAAAACGAGACATGACAACTAATGCTACTACAGGTGGTATTGTCAGC 2665
|||||

Search completed: May 30, 2006, 04:14:20
Job time : 44 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2006, 01:04:24 ; Search time 4485 Seconds
(without alignments)
8528.166 Million cell updates/sec

Title: US-10-718-311-7

Perfect score: 684

Sequence: 1 atggttctctctgtcatttc.....cggcgtcacccgttactaa 684

Scoring table: IDENTITY NUC

Gapop 10_0, Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est3:*
- 3: gb_est4:*
- 4: gb_est5:*
- 5: gb_est6:*
- 6: gb_htc:*
- 7: gb_est2:*
- 8: gb_est7:*
- 9: gb_est8:*
- 10: gb_est9:*
- 11: gb_gss1:*
- 12: gb_gss2:*
- 13: gb_gss3:*
- 14: gb_gss4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	363	53.1	832	13	CL666350 PRI0152b
2	295.4	43.2	828	13	CL687382
3	184	26.9	205	7	BE462847 EST325226
4	184	26.9	235	7	AW039464 EST281745
5	184	26.9	237	7	AW039041 EST281014
6	184	26.9	238	7	AW041024 EST283888
7	184	26.9	250	7	AW442951 EST307881
8	184	26.9	260	7	AW037699 EST279328
9	184	26.9	260	7	AW094122 EST287302
10	184	26.9	267	1	AI773928 EST255028
11	184	26.9	277	1	AI773502 EST254602
12	184	26.9	279	2	BI931188 EST551077
13	184	26.9	279	7	BE462486 EST324752
14	184	26.9	283	7	AW094091 EST287271
15	184	26.9	286	7	BF051382 EST436557
16	184	26.9	288	7	AW037811 EST279440
17	184	26.9	294	7	AW040243 EST282749
18	184	26.9	329	7	AW038470 EST280153
19	184	26.9	330	7	AW092013 EST285193

20	184	26.9	351	2	BG129648
21	184	26.9	356	2	BI927635
22	184	26.9	367	2	AW039845
23	184	26.9	370	1	AI773548
24	184	26.9	372	7	BE463004
25	184	26.9	373	7	AW093805
26	184	26.9	375	2	BG123527
27	184	26.9	377	7	AW038793
28	184	26.9	382	1	AI777051
29	184	26.9	385	7	AW092154
30	184	26.9	393	7	AW040522
31	184	26.9	394	2	BG128649
32	184	26.9	395	7	AW217063
33	184	26.9	401	7	AW093581
34	184	26.9	404	2	BM535307
35	184	26.9	407	2	BI929701
36	184	26.9	408	7	AW038880
37	184	26.9	412	2	BI928909
38	184	26.9	416	7	AW094716
39	184	26.9	417	7	AW933080
40	184	26.9	420	7	AW041064
41	184	26.9	420	7	AW443412
42	184	26.9	421	2	BG133589
43	184	26.9	421	2	BG642684
44	184	26.9	422	1	AI775340
45	184	26.9	423	7	AW041189

ALIGNMENTS

RESULT 1
CL666350/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

CL666350 832 bp DNA linear GSS 09-JUL-2004
PRI0152b_A03 - PRI0152b.B21 (832) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
CL666350
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 832)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.

FEATURES
Source

1..832
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

Query Match 53.1%; Score 363; DB 13; Length 832;

Best Local Similarity 100.0%; Pred. No. 1.8e-104; Mismatches 0; Gaps 0;

Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 GTAAGCGTGACGATGATCGCGAAGGGTTTGTGAGCAGAAATGAATCCCGAAGACTG 381

DB 832 GTAAGCGTGACGATGATCGCGAAGGGTTTGTGAGCAGAAATGAATCCCGAAGACTG 773

QY 382 CCGCTGCTGCCGAAAGAGTCTCGTTACTGCTTACGTGTTACGTGAAATTTTGTATGTCGCGATGT 441

DB 772 CCGCTGCTGCCGAAAGAGTCTCGTTACTGCTTACGTGTTACGTGAAATTTTGTATGTCGCGATGT 713

QY 442 GAACCGTGGCTTGCCTGCTACCGTCTGCTGCTCAACGTTAAGCGGGCCGAGCTG 501

DB 712 GAACCGTGGCTTGCCTGCTACCGTCTGCTGCTCAACGTTAAGCGGGCCGAGCTG 653

QY 502 GCGTTACAAAATTTGGGTAAACGCCGTTAGGACGCTATCTGTTACATCATCGACATTA 561

DB 652 GCGTTACAAAATTTGGGTAAACGCCGTTAGGACGCTATCTGTTACATCATCGACATTA 593

QY 562 ACCCGGGACTTTATTGAGATAGCGGTGATGCGGGCTGTGGGGGCGAGCTTCCGCGCTG 621

DB 592 ACCCGGGACTTTATTGAGATAGCGGTGATGCGGGCTGTGGGGGCGAGCTTCCGCGCTG 533

QY 622 CGATTAAAGCGGTAAACGCCGTTGCTTACAGAACTGTTTTTACCAGCGTCAACGTTGTAC 681

DB 532 CGATTAAAGCGGTAAACGCCGTTGCTTACAGAACTGTTTTTACCAGCGTCAACGTTGTAC 473

QY 682 TAA 684

DB 472 TAA 470

RESULT 2

LOCUS CL687382

DEFINITION PRI0146c_H03.2 - PRI0146c_BR (828) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL687382

VERSION CL687382.1 GI:50196135

KEYWORDS GSS.

SOURCE Pristionchus pacificus

ORGANISM Pristionchus pacificus

REFERENCE 1 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

TITLE Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J. AppaDB: an AcedB database for the nematode satellite organism

JOURNAL Pristionchus pacificus

PUBMED Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT 14681447

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: raif.sommer@tuebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.

Seq primer: T7

Class: fosmid ends.

Location/Qualifiers

1. :828

/organism="Pristionchus pacificus"

/mol_type="genomic DNA"

/strain="California"

/db_xref="taxon:54126"

/clone_lib="Mixed stage fosmid library of P. pacificus var. California"

/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 205;

Best Local Similarity 100.0%; Pred. No. 2.9e-47; Mismatches 0; Gaps 0;

Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCATTTCTTCAGCAGCTGTGTCACACGACGCAATGTTACACAAGCT 60

Query Match 43.2%; Score 295.4; DB 13; Length 828;

Best Local Similarity 99.7%; Pred. No. 8.7e-83; Mismatches 296; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 187 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 246

DB 532 ATGTCACACCCCGCTTAAACGCAACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 591

QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306

DB 592 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 651

QY 307 CAGCAGGGAAGAAACGGTTAAGCGTGACGATGATCCGCGAAGGTTTGTGAGCAGATGAA 366

DB 652 CAGCAGGGAAGAAACGGTTAAGCGTGACGATGATCCGCGAAGGTTTGTGAGCAGATGAA 711

QY 367 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGTGAAATTTG 426

DB 712 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGTGAAATTTG 771

QY 427 TTATGTGCCGATGTGAACCGTGTCCCGTTCGCGTTCGTACCGTTCCTGTGTCAACG 483

DB 772 TTATGTGCCGATGTGAACCGTGTCCCGTTCGCGTTCGTACCGTTCCTGTGTCAACG 828

RESULT 3

LOCUS BE462847

DEFINITION EST325226 tomato flower buds 0-3 mm, Cornell University Lycopersicon esculentum cDNA clone cTOA16011, mRNA sequence.

ACCESSION BE462847

VERSION BE462847.1 GI:9508618

KEYWORDS EST.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 205)

AUTHORS Liang,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, 0-3 mm buds

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1. :205

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cTOA16011"

/tissue_type="flower"

/dev_stage="0-3mm buds"

/clone_lib="tomato flower buds 0-3 mm, Cornell University"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

FEATURES

source

1. :205

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cTOA16011"

/tissue_type="flower"

/dev_stage="0-3mm buds"

/clone_lib="tomato flower buds 0-3 mm, Cornell University"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 205;

Best Local Similarity 100.0%; Pred. No. 2.9e-47; Mismatches 0; Gaps 0;

Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCATTTCTTCAGCAGCTGTGTCACACGACGCAATGTTACACAAGCT 60

Db 8 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGCAATGTTACACAAGCT 67
 QY 61 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCACTTTCCTGTTACAAGAAG 120
 Db 68 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCACTTTCCTGTTACAAGAAG 127
 QY 121 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180
 Db 128 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 187
 QY 181 TGGC 184
 Db 188 TGGC 191

RESULT 4
 AW039464
 LOCUS
 DEFINITION EST281745 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
 ACCESSION AW039464
 VERSION AW039464.1 GI:5898218
 KEYWORDS EST.
 ORGANISM Lycopersicon esculentum (Solanum lycopersicum)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 235)
 REFERENCE
 AUTHORS D'Ascenzo, M., He, X., Lyman, J., Holt, I. E., Liang, F., Upton, J.,
 Ronning, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Nierman, W.,
 Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D. and
 Giovannoni, J.
 TITLE Generation of ESTs from tomato leaf tissue
 JOURNAL Unpublished (1999)
 COMMENT Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source
 1..235
 Location/Qualifiers
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="Rio Grande PtoR"
 /db_xref="taxon:4081"
 /clone="cLET10112"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="XL1-Blue MRF"
 /clone_lib="tomato mixed elicitor, BTI"
 /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLET - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisonicotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."

ORIGIN
 Query Match 26.9%; Score 184; DB 7; Length 235;
 Best Local Similarity 100.0%; Pred. No. 3e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGCAATGTTACACAAGCT 60
 Db 14 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGCAATGTTACACAAGCT 73
 QY 61 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCACTTTCCTGTTACAAGAAG 120
 Db 74 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCACTTTCCTGTTACAAGAAG 133

QY 121 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180
 Db 134 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 193
 QY 181 TGGC 184
 Db 194 TGGC 197

RESULT 5
 AW039041
 LOCUS
 DEFINITION EST281014 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
 ACCESSION AW039041
 VERSION AW039041.1 GI:5897795
 KEYWORDS EST.
 ORGANISM Lycopersicon esculentum (Solanum lycopersicum)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 237)
 REFERENCE
 AUTHORS D'Ascenzo, M., He, X., Lyman, J., Holt, I. E., Liang, F., Upton, J.,
 Ronning, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Nierman, W.,
 Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D. and
 Giovannoni, J.
 TITLE Generation of ESTs from tomato leaf tissue
 JOURNAL Unpublished (1999)
 COMMENT Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source
 1..237
 Location/Qualifiers
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="Rio Grande PtoR"
 /db_xref="taxon:4081"
 /clone="cLET10C2"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="XL1-Blue MRF"
 /clone_lib="tomato mixed elicitor, BTI"
 /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLET - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisonicotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."

ORIGIN
 Query Match 26.9%; Score 184; DB 7; Length 237;
 Best Local Similarity 100.0%; Pred. No. 3e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGCAATGTTACACAAGCT 60
 Db 9 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGCAATGTTACACAAGCT 68
 QY 61 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCACTTTCCTGTTACAAGAAG 120
 Db 69 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCACTTTCCTGTTACAAGAAG 128
 QY 121 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180
 Db 129 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 188
 QY 181 TGGC 184
 Db 189 TGGC 192

```

RESULT 6
AW041024
LOCUS
DEFINITION
  EST283888 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
  clone cLET615, mRNA sequence.
ACCESSION
AW041024
VERSION
AW041024.1
KEYWORDS
GI:5899778
SOURCE
  Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 238)
AUTHORS
  D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
  Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
  Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and
  Giovannoni,J.
TITLE
  Generation of ESTs from tomato leaf tissue
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: CUGI
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
FEATURES
  source
  1..238
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="Rio Grande PtoR"
    /db_xref="taxon:4081"
    /clone="cLET615"
    /tissue_type="leaf"
    /dev_stage="4-6 week old plants"
    /lab_host="XLI-Blue MRF"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; cLET - Inoculated with a variety of disease response
    elicitors. Plants exposed to 2,6 dichloroisonicotinic
    acid, BTH, jasmonic acid, ethylene, fenthion, BIX,
    okadaic acid, or systemin prior to tissue harvest. EcoRI
    site was destroyed during cloning."
ORIGIN
  Query Match      26.9%; Score 184; DB 7; Length 238;
  Best Local Similarity 100.0%; Pred. No. 3e-47;
  Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db 19 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 78

QY 61 AGCATGGTTGCACCTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 120
Db 79 AGCATGGTTGCACCTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 138

QY 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 180
Db 139 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 198

QY 181 TGGC 184
Db 199 TGGC 202

RESULT 7
AW442951
LOCUS
DEFINITION
  EST307881 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
  clone cLET42H11 5', mRNA sequence.
ACCESSION
AW442951
VERSION
AW442951.1
KEYWORDS
GI:5895133
SOURCE
  Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 250)
AUTHORS
  D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
  Liang,F., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
  Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
  Giovannoni,J.J. and Martin,G.B.
TITLE
  Generation of ESTs from tomato callus (mixed elicitor)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: CUGI
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
FEATURES
  source
  1..250
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="Rio Grande PtoR"
    /db_xref="taxon:4081"
    /clone="cLET42H11"
    /tissue_type="leaf"
    /dev_stage="4-6 week old plants"
    /lab_host="XLI-Blue MRF"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; cLET - Inoculated with a variety of disease response
    elicitors. Plants exposed to 2,6 dichloroisonicotinic
    acid, BTH, jasmonic acid, ethylene, fenthion, BIX,
    okadaic acid, or systemin prior to tissue harvest. EcoRI
    site was destroyed during cloning."
ORIGIN
  Query Match      26.9%; Score 184; DB 7; Length 250;
  Best Local Similarity 100.0%; Pred. No. 3.1e-47;
  Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db 6 ATGGCTTCTCTGTCATTTCTTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 65

QY 61 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 120
Db 66 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 125

QY 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 180
Db 126 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 185

QY 181 TGGC 184
Db 186 TGGC 189

RESULT 8
AW037699
LOCUS
DEFINITION
  EST279328 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
  clone cLET4M17, mRNA sequence.
ACCESSION
AW037699
VERSION
AW037699.1
KEYWORDS
GI:5896453
SOURCE
  Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

```

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 260)
 D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
 Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Niernan, W.,
 Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and
 Giovannoni, J.
 Generation of ESTs from tomato leaf tissue
 Unpublished (1999)
 CONTACT: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source
 1..260
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="Rio Grande PtoR"
 /db_xref="taxon:4081"
 /clone="cLET4M17"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="XLI-Blue MRF"
 /clone_lib="tomato mixed elicitor, BTI"
 /notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLET - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisocotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, EXX, EcoRI
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."

ORIGIN
 Query Match 26.9%; Score 184; DB 7; Length 260;
 Best Local Similarity 100.0%; Pred. No. 3.1e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACAGCGCAATGTTACACAAGCT 60
 |||||
 Db 12 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACAGCGCAATGTTACACAAGCT 71
 |||||

QY 61 AGCATGGTGGACCTTTCATCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
 |||||
 Db 72 AGCATGGTGGACCTTTCATCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 131
 |||||

QY 121 CAAACCTTGACATCACTTCCATCTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180
 |||||
 Db 132 CAAACCTTGACATCACTTCCATCTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 191
 |||||

QY 181 TGGC 184
 |||||
 Db 192 TGGC 195

RESULT 9
 AW094122
 LOCUS
 DEFINITION
 EST287302 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
 clone cLET27G20, mRNA sequence.
 ACCESSION
 AW094122
 VERSION
 AW094122.1 GI:6059717
 EST.
 SOURCE
 Lycopersicon esculentum (Solanum lycopersicum)
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 260)
 D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
 Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Niernan, W.,
 Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and
 Giovannoni, J.
 Generation of ESTs from tomato leaf tissue

Unpublished (1999)
 CONTACT: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 Location/Qualifiers
 1..260
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="Rio Grande PtoR"
 /db_xref="taxon:4081"
 /clone="cLET27G20"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="XLI-Blue MRF"
 /clone_lib="tomato mixed elicitor, BTI"
 /notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLET - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisocotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, EXX,
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."

ORIGIN
 Query Match 26.9%; Score 184; DB 7; Length 260;
 Best Local Similarity 100.0%; Pred. No. 3.1e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACAGCGCAATGTTACACAAGCT 60
 |||||
 Db 6 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACAGCGCAATGTTACACAAGCT 65
 |||||

QY 61 AGCATGGTGGACCTTTCATCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
 |||||
 Db 66 AGCATGGTGGACCTTTCATCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 125
 |||||

QY 121 CAAACCTTGACATCACTTCCATCTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180
 |||||
 Db 126 CAAACCTTGACATCACTTCCATCTAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 185
 |||||

QY 181 TGGC 184
 |||||
 Db 186 TGGC 189

RESULT 10
 AI773928
 LOCUS
 DEFINITION
 EST255028 tomato resistant, Cornell Lycopersicon esculentum cDNA
 clone cLER8L7, mRNA sequence.
 ACCESSION
 AI773928
 VERSION
 AI773928.1 GI:5271969
 EST.
 KEYWORDS
 SOURCE
 Lycopersicon esculentum (Solanum lycopersicum)
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 267)
 D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
 Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y.,
 Bowman, C.L., Niernan, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
 Giovannoni, J. and Martin, G.B.
 Generation of ESTs from Pseudomonas resistant tomato
 Unpublished (1999)
 CONTACT: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

```

FEATURES
  source
    Location/Qualifiers
      1..267
        /organism="Lycopersicon esculentum"
        /mol_type="mRNA"
        /cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
        /db_xref="taxon:4081"
        /clone="cLER8L7"
        /tissue_type="leaf"
        /dev_stage="4-week old"
        /lab_host="SOLR"
        /clone_lib="tomato resistant, Cornell"
        /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; cLER - Tomato Pseudomonas Resistant EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN
  Query Match      26.9%; Score 184; DB 1; Length 267;
  Best Local Similarity 100.0%; Pred. No. 3.2e-47;
  Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATGGCTTCTGTCATTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db      23  ATGGCTTCTGTCATTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 82

QY      61  AGCATGGTGGACCTTCACTGGTCTCAAACTCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
Db      83  AGCATGGTGGACCTTCACTGGTCTCAAACTCTTCAGCCACTTTCCTGTTTACAAAGAAG 142

QY      121  CAAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAAGTG 180
Db      143  CAAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAAGTG 202

QY      181  TGGC 184
Db      203  TGGC 206

RESULT 11
A1773502
LOCUS      A1773502      277 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION EST254602 tomato resistant, Cornell Lycopersicon esculentum cDNA clone cLER7A10, mRNA sequence.
ACCESSION A1773502
VERSION   A1773502.1 GI:5271543
KEYWORDS EST.
SOURCE    Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
    asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
  1 (bases 1 to 277)
    D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
    Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y.,
    Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
    Giovannoni, J.J. and Martin, G.B.
    Generation of ESTs from Pseudomonas resistant tomato
  Unpublished (1999)
  Contact: CUGI
  Institution: Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
  Location/Qualifiers
    1..277
      /organism="Lycopersicon esculentum"
      /mol_type="mRNA"
      /cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
      /db_xref="taxon:4081"
      /clone="cLER7A10"
      /tissue_type="leaf"
      /dev_stage="4-week old"

FEATURES
  source
    Location/Qualifiers
      1..279
        /organism="Lycopersicon esculentum"
        /mol_type="mRNA"
        /cultivar="TA496"
        /db_xref="taxon:4081"
        /clone="cTOC19C6"
        /tissue_type="flower"
        /dev_stage="buds 8mm to preanthesis"
        /clone_lib="tomato flower, 8 mm to preanthesis buds"
        /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

```

```

/lab_host="SOLR"
/clone_lib="tomato resistant, Cornell"
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; cLER - Tomato Pseudomonas Resistant EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN
  Query Match      26.9%; Score 184; DB 1; Length 277;
  Best Local Similarity 100.0%; Pred. No. 3.2e-47;
  Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATGGCTTCTGTCATTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db      5  ATGGCTTCTGTCATTCTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 64

QY      61  AGCATGGTGGACCTTTCACCTGGTCTCAAACTCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
Db      65  AGCATGGTGGACCTTTCACCTGGTCTCAAACTCTTCAGCCACTTTCCTGTTTACAAAGAAG 124

QY      121  CAAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAAGTG 180
Db      125  CAAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAAGTG 184

QY      181  TGGC 184
Db      185  TGGC 188

RESULT 12
BI931188
LOCUS      BI931188      279 bp      mRNA      linear      EST 18-OCT-2001
DEFINITION EST551077 tomato flower, 8 mm to preanthesis buds Lycopersicon esculentum cDNA clone cTOC19C6 5' end, mRNA sequence.
ACCESSION BI931188
VERSION   BI931188.1 GI:16245660
KEYWORDS EST.
SOURCE    Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
    asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
  1 (bases 1 to 279)
    van der Hoeven, R.S., Bezzerides, J.L., Karanycheva, S.A., Tsai, J.,
    Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
    Generation of ESTs from tomato flower tissue, buds 8 mm -
    preanthesis
  Unpublished (2001)
  Contact: CUGI
  Institution: Clemson University Genomics Institute
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  This clone is available through the Clemson University Genomics
  Institute
  Seq primer: T3.
  Location/Qualifiers
    1..279
      /organism="Lycopersicon esculentum"
      /mol_type="mRNA"
      /cultivar="TA496"
      /db_xref="taxon:4081"
      /clone="cTOC19C6"
      /tissue_type="flower"
      /dev_stage="buds 8mm to preanthesis"
      /clone_lib="tomato flower, 8 mm to preanthesis buds"
      /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

```

ORIGIN

Query Match 26.9%; Score 184; DB 2; Length 279;
 Best Local Similarity 100.0%; Pred. No. 3.2e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAGCT 60
 |||||
 Db 7 ATGGCTTCCTGTCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAGCT 66

QY 61 AGCATGGTTCACCTTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 120
 |||||
 Db 67 AGCATGGTTCACCTTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 126

QY 121 CAAACCTTGACATCTTCCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 180
 |||||
 Db 127 CAAACCTTGACATCTTCCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 186

QY 181 TGGC 184
 |||||
 Db 187 TGGC 190

RESULT 13

BE462486 279 bp mRNA linear EST 18-MAY-2001
 EST324752 tomato flower buds 0-3 mm, Cornell University
 Lycopersicon esculentum cDNA clone CTOA1301, mRNA sequence.

ACCESSION BE462486
 VERSION
 KEYWORDS

SOURCE EST.
 ORGANISM Lycopersicon esculentum (Solanum lycopersicum)

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 279)
 van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E.,
 Liang, F., Hansen, T., Craven, M.B., Bowman, C.L., Ronning, C.M.,
 Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and
 Tanksley, S.D.

TITLE Generation of ESTs from tomato flower tissue, 0-3 mm buds
 JOURNAL Unpublished (1999)

COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES

source Location/Qualifiers
 1..279
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="CTOA1301"
 /tissue_type="flower"
 /dev_stage="0-3 mm buds"
 /clone_lib="tomato flower buds 0-3 mm, Cornell University"
 /notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Tanksley; Flower buds and flowers were
 taken from greenhouse plants (4-8 wks old, TA496). They
 were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."

ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 279;
 Best Local Similarity 100.0%; Pred. No. 3.2e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAGCT 60
 |||||
 Db 7 ATGGCTTCCTGTCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAGCT 66

QY 61 AGCATGGTTCACCTTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 120
 |||||
 Db 67 AGCATGGTTCACCTTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 126

QY 121 CAAACCTTGACATCTTCCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 180
 |||||
 Db 127 CAAACCTTGACATCTTCCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 186

QY 181 TGGC 184
 |||||
 Db 187 TGGC 190

RESULT 14

AW094091 283 bp mRNA linear EST 18-MAY-2001
 EST287271 tomato mixed elicitor, B71 Lycopersicon esculentum cDNA
 clone cLET27A10, mRNA sequence.

ACCESSION AW094091
 VERSION
 KEYWORDS

SOURCE EST.
 ORGANISM Lycopersicon esculentum (Solanum lycopersicum)

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 283)
 D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
 Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,
 Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and
 Giovannoni, J.

TITLE Generation of ESTs from tomato leaf tissue
 JOURNAL Unpublished (1999)

COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES

source Location/Qualifiers
 1..283
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="Rio Grande Ptor"
 /db_xref="taxon:4081"
 /clone="cLET27A10"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="XLI-Blue MRF"
 /clone_lib="tomato mixed elicitor, BTI"
 /notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLET - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisonicotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, Etx,
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."

ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 283;
 Best Local Similarity 100.0%; Pred. No. 3.2e-47;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAGCT 60
 |||||
 Db 12 ATGGCTTCCTGTCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAGCT 71

QY 61 AGCATGGTTCACCTTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 120
 |||||
 Db 72 AGCATGGTTCACCTTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 131

QY 121 CAAACCTTGACATCTTCCATTTCTTCCAGCAGCTGTTGCCACAGCAAGTTCACAAAG 180
 |||||

```

Db      132 CAAACCTTGACATCACTTCATTGCTAGCAATGGTGGAAAGATTAGTGCATGCAGGTG 191
Qy      181 TGGC 184
        ||||
Db      192 TGGC 195

RESULT 15
BF051382
LOCUS   EST436557 tomato developing/immature green fruit Lycopersicon
DEFINITION
BF051382 esculentum cDNA clone CLEM22P19 5' sequence, mRNA sequence.
ACCESSION
BF051382
VERSION
BF051382.1 GI:10805278
KEYWORDS
EST.
SOURCE
Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 286)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and
Tanksey,S.D.
Generation of ESTs from tomato fruit tissue, immature green
Unpublished (2000)
Contact: CUGI
Clemson University
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1..286
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEM22P19"
/tissue_type="fruit"
/dev_stage="immature green (5-35 days post-anthesis)"
/lab_host="SOLR"
/clone_lib="tomato developing/immature green fruit"
/notes="Vector: pBluescriptSKmCtadapt; Site1: EcoRI;
Site 2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
harvested at 7 day intervals through 35 dpa. Equal masses
of tissue from each stage were combined (including seeds
and locules) prior to mRNA isolation."

ORIGIN
Query Match 26.9%; Score 184; DB 7; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.3e-47;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ATGGCTTCCTCTGTCATTCTTCACGAGCTGTTCACACGACGCAATGTTACACAAGCT 60
Db      10 ATGGCTTCCTCTGTCATTCTTCACGAGCTGTTCACACGACGCAATGTTACACAAGCT 69

Qy      61 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db      70 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 129

Qy      121 CAAACCTTGACATCACTTCATTCCTAGCAATGCTGGAAGATTAGCTGCATGCAGGTG 180
Db      130 CAAACCTTGACATCACTTCATTCCTAGCAATGCTGGAAGATTAGCTGCATGCAGGTG 189

Qy      181 TGGC 184
        ||||
Db      190 TGGC 193

```

Search completed: May 30, 2006, 03:27:40
Job time : 4488 secs

GenCore version 5.1.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2006, 01:30:02 ; Search time 5152.27 Seconds

(without alignments)
3164.934 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVHMSHPALQRLALRYC.....RLSGKPLLLTELFLPASPLY 170

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame*_p2n.model -DRV=xlh
-Q=/abs/ABSSWEB_spool/US10718311/runat_26052006_164916_13439/app_query.fasta_1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs03h
-USER=US10718311@CGN_1_5767@runat_26052006_164916_13439 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

1: gb_env.*
2: gb_pat.*
3: gb_ph.*
4: gb_pl.*
5: gb_pr.*
6: gb_ro.*
7: gb_sts.*
8: gb_sy.*
9: gb_un.*
10: gb_vi.*
11: gb_ov.*
12: gb_htg.*
13: gb_in.*
14: gb_om.*
15: gb_ba.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	887	100.0	684	2	AR428736 Sequence
2	887	100.0	684	2	AR455412 Sequence
3	887	100.0	684	2	AX329368 Sequence

4	887	100.0	684	2	AX382258	AX382258 Sequence
5	854	96.3	495	2	A47932	A47932 Sequence 1
6	854	96.3	495	2	AR428733	AR428733 Sequence
7	854	96.3	495	2	AR455409	AR455409 Sequence
8	854	96.3	495	2	AX329364	AX329364 Sequence
9	854	96.3	495	2	AX382254	AX382254 Sequence
10	854	96.3	854	2	AR144788	AR144788 Sequence
11	854	96.3	1498	15	ECUBIFA	M96268 Escherichia
12	854	96.3	2000	2	E11274	E11274 ubiC and ub
13	854	96.3	2000	15	ECUBICA	M93136 Escherichia
14	854	96.3	2000	15	ECUBIPLS	M93413 Escherichia
15	854	96.3	2034	15	ECUBIAC	X57434 E.coli ubiC
16	854	96.3	2348	15	ECUBI	X66619 E.coli gene
17	854	96.3	6641	2	AR144789	AR144789 Sequence
18	854	96.3	110000	15	U00096_42	Continuation (43 o
19	854	96.3	110000	15	AP009048_42	Continuation (43 o
20	854	96.3	176195	15	ECOUW89	U00006 E. coli chr
21	850	95.8	110000	15	AE005174_51	Continuation (52 o
22	846	95.4	110000	15	BA000007_50	Continuation (51 o
23	846	95.4	110000	15	BA000007_51	Continuation (52 o
24	843	95.0	1396	15	DQ087228	DQ087228 Escherich
25	843	95.0	110000	15	AE005674_43	Continuation (44 o
26	843	95.0	110000	15	AE014073_34	Continuation (35 o
27	843	95.0	110000	15	CP000036_41	Continuation (42 o
28	839	94.6	110000	15	CP000036_42	Continuation (43 o
29	835	94.1	110000	15	CP000038_44	Continuation (45 o
30	833	93.9	110000	15	AE014075_47	Continuation (48 o
31	798	90.0	1318	2	E07849	E07849 gDNA encodi
32	725	81.7	23880	15	AE008898	AE008898 Salmonell
33	725	81.7	110000	15	CP000026_41	Continuation (42 o
34	725	81.7	110000	15	CP000026_42	Continuation (43 o
35	721	81.3	110000	15	AE014613_42	Continuation (43 o
36	721	81.3	110000	15	AE017220_43	Continuation (44 o
37	721	81.3	234050	15	AL627282	AL627282 salmonell
38	673	75.9	645	2	AR384505	AR384505 Sequence
39	437.5	49.3	10668	15	AE013658	AE013658 Versinia
40	437.5	49.3	110000	15	EX936398_04	Continuation (5 of
41	437.5	49.3	214050	15	AJ414142	AJ414142 Versinia
42	437.5	49.3	290803	15	AE017128	AE017128 Versinia
43	424.5	47.9	110000	15	EX950851_06	Continuation (7 of
44	417	47.0	349107	15	EX571873	EX571873 Phototrab
45	408.5	46.1	110000	15	EX897699_12	Continuation (13 o

ALIGNMENTS

RESULT 1	AR428736	Sequence 16 from patent US 6642036.	684 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR428736	Sequence 16 from patent US 6642036.				
DEFINITION	AR428736					
ACCESSION	AR428736.1	GI:40188466				
VERSION	AR428736.1	GI:40188466				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 684)					
AUTHORS	Flint, D., Meyer, K. and Viitanen, P.					
TITLE	Sinapoylgucose:malate sinapoyltransferase form malate conjugates from benozic acid glucosides					
JOURNAL	Patent: US 6642036-A 16 04-NOV-2003;					
FEATURES	WOX;					
source	Location/Qualifiers					
	1..684					
	/organism="unknown"					
	/mol_type="genomic DNA"					
ORIGIN						
Alignment Scores:						
Pred. No.:	1.23e-83					684
Score:	887.00					170
Percent Similarity:	100.0%					Conservative: 0

```

Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR428736 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTCACACCCCGCGTTAACGCAACTGCGTGGCGCTATTGTTGT 231

Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTGCACCTGGCTGTTGCTGGAGGATCCATG 291

Qy 41 ThrLysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAACGTTTTTGAACACGACGAGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGGTTT 351

Qy 61 ValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGAGTAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411

Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAATTTTGTATGTGCGCATGGTGAACCGTGGCTTGGCGGTCTGACCGTCTGTT 471

Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACGCCGTTA 531

Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTTACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGAT 591

Qy 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 3
AX329368
LOCUS AX329368 684 bp DNA linear PAT 08-JAN-2002
DEFINITION Sequence 7 from Patent WO0194607.
ACCESSION AX329368
VERSION AX329368.1 GI:18102367
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Meyer, K., van Dyk, D.E. and Viitanen, P.V.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: WO 0194607-A 7 13-DEC-2001;
FEATURES
Location/Qualifiers
source
1. .684
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="synthetic CPL"

ORIGIN
Alignment Scores:
Pred. No.: 1.23e-83 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AX329368 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTCACACCCCGCGTTAACGCAACTGCGTGGCGCTATTGTTGT 231

Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTGCACCTGGCTGTTGCTGGAGGATCCATG 291

Qy 41 ThrLysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAACGTTTTTGAACACGACGAGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGGTTT 351

Qy 61 ValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGAGTAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411

Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAATTTTGTATGTGCGCATGGTGAACCGTGGCTTGGCGGTCTGACCGTCTGTT 471

Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACGCCGTTA 531

Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTTACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGAT 591

Qy 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 2
AR455412
LOCUS AR455412 684 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 7 from patent US 6683231.
ACCESSION AR455412
VERSION AR455412.1 GI:42689960
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 684)
AUTHORS Meyer, K., Viitanen, P.V. and Van Dyk, D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 7 27-JAN-2004;
FEATURES
Location/Qualifiers
source
1. .684
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1.23e-83 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR455412 (1-684)

```

Db	232	AAAGAGATCCCTGCGCTCGATCCGCAACTGCTGCAGCTGGCTGTGCTGGAGGATTCATG	291
Qy	41	ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe	60
Db	292	ACAAAACGTTTTGAACACACGCGAAAAACGGTAAACGGTACCGATGATCCGCGAAGGGTTT	351
Qy	61	ValGluGlnAanGluIleProGluGluLeuProLeuProLysGluSerArgTyrTrp	80
Db	352	GTCGAGCAAGATGAATCCCCGGAAGACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGG	411
Qy	81	LeuArgGluIleLeuLeuCyAlaAspGlyGluProTrrLeuAlaGlyVargThrValVal	100
Db	412	TTACGTGAAATTTTGTTATGTGCCGATGGTGNACCGTGGCTTGCCTGCTACCGTCGTT	471
Qy	101	ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu	120
Db	472	CCTGTGTCAACGTTAAGCGGGCCGAGCTGGCGTTTACAAAAATTGGGTAAAAACGCCGTTA	531
Qy	121	GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeGlyArgAsp	140
Db	532	GGACGCTATCTGTTACATCATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGAT	591
Qy	141	AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr	160
Db	592	GCCGGCTGTGGGGCGACGTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTGTCTAACA	651
Qy	161	GluLeuPheLeuProAlaSerProLeuTyr	170
Db	652	GAACTGTTTTTACCGGGGTCACCGTTGTAC	681

A47932	A47932	495 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	Sequence 1 from Patent WO9600788.				
DEFINITION	A47932				
ACCESSION	A47932.1	GI:2301795			
VERSION					
KEYWORDS	Escherichia coli				
SOURCE	Escherichia coli				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
	Enterobacteriaceae; Escherichia.				
REFERENCE	1. (bases 1 to 495)				

FEATURES	source	ORIGIN
COMMENTS	Other publication AU 2879295 960125 Other publication DE 4423022 950524. Location/Qualifiers	
FEATURES	1. .495 /organism="Escherichia coli" /mol_type="unassigned DNA" /strain="MC4100" /db_xref="taxon:562"	
COMMENTS	Heide, L., Severin, K. and Siebert, M. TRANSGENIC PLANTS WITH AN INCREASED PATENT: WO 9600788-A 11-JAN-1996; <u>HEIDE-0072 (DS)</u>	
FEATURES	1. 1033 /organism="Escherichia coli" /mol_type="unassigned DNA" /strain="MC4100" /db_xref="taxon:562"	

ORIGIN	
Alignment Scores:	
Pred. No.:	2,59e-80
Score:	854.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	96.3%
DB:	2
	Length: 495
	Matches: 165
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0

US-10-/18-311-16 (1-170) X A47932 (1-495).

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTACACCCCGTTAACGCAACTGCCTGCGCTATTGTAAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrlsArgPheGlu	45
Db	61	CTGGATCCGCAACTGTCACATGCGTGTTGCTGGAGATTTCATGACAAAAGCTTTGAA	120

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 121 CAGCAGGGGAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTTGTTCGACGAGATGAA 180

Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 181 ATCCCGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTCGTTACGTTAAATTTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTACAAAATTTGGTAAACGCCGTTAGGACGTAATC 300

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTACAAAATTTGGTAAACGCCGTTAGGACGTAATC 300

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIlePhe 125
Db 301 AGCGGGCGGAGCTGGCGTTTACAAAATTTGGTAAACGCCGTTAGGACGTAATC 360

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAACCCGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGAGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTTACCG 480

Qy 166 AlaSerProLeuTyr 170
Db 481 CGGTACCGTTGTAC 495

RESULT 6
AR428733
LOCUS AR428733 495 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 12 from patent US 6642036.
ACCESSION AR428733
VERSION AR428733.1 GI:40188463
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Flint, D., Meyer, K. and Viitanen, P.
TITLE Sinapoylglucose:malate sinapoyltransferase form malate conjugates from benzoic acid glucosides
JOURNAL Patent: US 6642036-A 12 04-NOV-2003;
E. I. Du Pont de Nemours and Company; Wilmington;
WOX;

FEATURES
source Location/Qualifiers
1. .495
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 2.59e-80 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR428733 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 1 ATGTCACACCCCGCGTTAAACGCACTGCGCTGCGCTATTGTAAAGAGATCCCTGCC 60

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGATCCGCACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 121 CAGCAGGGGAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTTGTTCGACGAGATGAA 180

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db 181 ATCCCGAAGAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTTGTTCGACGAGATGAA 180

Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 181 ATCCCGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTCGTTAAATTTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTACAAAATTTGGTAAACGCCGTTAGGACGTAATC 300

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIlePhe 125
Db 301 AGCGGGCGGAGCTGGCGTTTACAAAATTTGGTAAACGCCGTTAGGACGTAATC 360

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAACCCGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGAGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTAACAGAACTGTTTTTACCG 480

Qy 166 AlaSerProLeuTyr 170
Db 481 CGGTACCGTTGTAC 495

RESULT 7
AR455409
LOCUS AR455409 495 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 3 from patent US 6683231.
ACCESSION AR455409
VERSION AR455409.1 GI:42689957
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Meyer, K., Viitanen, P. V. and Van Dyk, D. E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 3 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE

FEATURES
source Location/Qualifiers
1. .495
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 2.59e-80 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR455409 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 1 ATGTCACACCCCGCGTTAAACGCACTGCGCTGCGCTATTGTAAAGAGATCCCTGCC 60

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGATCCGCACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 121 CAGCAGGGGAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTTGTTCGACGAGATGAA 180

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db 181 ATCCCGAAGAAACCGTTCGCTGCTCCGAAAGAGTCTCGTTACTCGTTAAATTTTG 240

```
Db      241 TTATGTGCCGATGGTGAACCGTGGCTTGGCGTGCATACCGTCTCTGTGTCAACGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db      301 AGCGGGCCGAGTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy      126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTATCCCGGGCTGTGGGG 420
Qy      146 ArgArgSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTATCCCGGGCTGTGGGG 420
Qy      146 ArgArgSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      421 CGACGTTCCCGCTCGATTAAACCGGTAAACCGCTGTTCGTAACGAACTGTTTTACCG 480
Qy      166 AlaSerProLeuTyr 170
Db      481 GCGTCACCGTTGTAC 495

RESULT 8
AX329364
LOCUS      AX329364
DEFINITION Sequence 3 from Patent WO0194607.
ACCESSION AX329364
VERSION    AX329364.1 GI:18102364
KEYWORDS   .
SOURCE     unidentified
ORGANISM   unidentified
REFERENCE  1
AUTHORS    Meyer, K., van Dyk, D. E. and Viitanen, P. V.
TITLE      High level production of p-hydroxybenzoic acid in green plants
JOURNAL    Patent: WO 0194607-A 3 13-DEC-2001;
           E. I. Dupont De Nemours (US)
FEATURES   Location/Qualifiers
            source
              1..495
              /organism="unidentified"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32644"
              /note="Description of Unknown Organism:E. coli"

ORIGIN
Alignment Scores:
Pred. No.:      2,596-80      Length:      495
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    96.3%      Indels:      0
DB:             2          Gaps:        0

US-10-718-311-16 (1-170) x AX329364 (1-495)

Qy      6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db      1 ATGTCACACCCCGCTTAACGCAACTGGCTGGCGCTATTGTAAAGAGATCCCTGCC 60
Qy      26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db      61 CTGGATCCGCAACTGCTCGACTGGCTGTCTGGAGGATTCATGACAAAAGCTTTTGA 120
Qy      46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65
Db      121 CAGCAGGGAAAACCGTAAGCGTGACGATGATCCGGAAGGGTTTGTGAGCAGAAATGAA 180
Qy      66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db      181 ATCCCGGAAGAACTCCCGCTGTCGCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 240
Qy      86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db      241 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTCTGACCGTCTCTGTGTCAACGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
```

```
Db      301 AGCGGGCCGAGTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy      126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTATCCCGGGCTGTGGGG 420
Qy      146 ArgArgSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      421 CGACGTTCCCGCTCGATTAAACCGGTAAACCGCTGTTCGTAACGAACTGTTTTACCG 480
Qy      166 AlaSerProLeuTyr 170
Db      481 GCGTCACCGTTGTAC 495

RESULT 9
AX382254
LOCUS      AX382254
DEFINITION Sequence 12 from Patent WO0204653.
ACCESSION AX382254
VERSION    AX382254.1 GI:19577034
KEYWORDS   .
SOURCE     Escherichia coli
ORGANISM   Escherichia coli
REFERENCE  1
AUTHORS    Flint, D., Meyer, K. and Viitanen, P. V.
TITLE      Sinapolyglucose:maltate sinapolytransferase form malate conjugates
           from benzoic acid glucosides
JOURNAL    Patent: WO 0204653-A 12 17-JAN-2002;
           E. I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES   Location/Qualifiers
            source
              1..495
              /organism="Escherichia coli"
              /mol_type="unassigned DNA"
              /db_xref="taxon:562"

ORIGIN
Alignment Scores:
Pred. No.:      2,596-80      Length:      495
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    96.3%      Indels:      0
DB:             2          Gaps:        0

US-10-718-311-16 (1-170) x AX382254 (1-495)

Qy      6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db      1 ATGTCACACCCCGCTTAACGCAACTGGCTGGCGCTATTGTAAAGAGATCCCTGCC 60
Qy      26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db      61 CTGGATCCGCAACTGCTCGACTGGCTGTCTGGAGGATTCATGACAAAAGCTTTTGA 120
Qy      46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65
Db      121 CAGCAGGGAAAACCGTAAGCGTGACGATGATCCGGAAGGGTTTGTGAGCAGAAATGAA 180
Qy      66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db      181 ATCCCGGAAGAACTCCCGCTGTCGCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 240
Qy      86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db      241 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTCTGACCGTCTCTGTGTCAACGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db      301 AGCGGGCCGAGTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
```

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
 Db 361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTGTGCGGG 420

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 Db 421 CGAGTTCCCGCTGGCGATTAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 480

Qy 166 AlaSerProLeuTyr 170
 Db 481 CGGTACCGTTGTAC 495

RESULT 10
 AR144788
 LOCUS
 DEFINITION Sequence 24 from patent US 6210937.
 ACCESSION AR144788
 VERSION AR144788.1 GI:15106655
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE
 1 (bases 1 to 854)
 AUTHORS Ward, T. E., Watkins, C. S., Bulmer, D. K., Johnson, B. F. and Amarantunga, M.
 TITLE Development of genetically engineered bacteria for production of selected aromatic compounds
 JOURNAL Patent: US 6210937-A 24 03-APR-2001;
 FEATURES Location/Qualifiers
 source 1..854
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 4,67e-80 Length: 854
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR144788 (1-854)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
 Db 321 ATGTCACACCCCGGTTAAACGAACTGCGTGGCTGCCTATTGTAAAGAGATCCCTGCC 380

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
 Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAAACGTTTGAA 440

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
 Db 441 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGTTGTTCGACAGAAATGAA 500

Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
 Db 501 ATCCCGGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGTTACGTTGAAATTTTG 560

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrIleu 105
 Db 561 TTATGTGCGCATGTTGGTAACCGTGGCTTCCCGGTCGTACCGTCGTTCCTGTGTCAACGTTA 620

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuPhe 125
 Db 621 ACGGGCCGAGACTGGCGTTACAAAATTGGTAAACCGCGTAGGACGCTATCTGTTTC 680

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
 Db 681 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTGTGCGGGCTGTGCGGG 740

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165

Db 741 CGAGTTCCCGCTGGCGATTAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 800

Qy 166 AlaSerProLeuTyr 170
 Db 801 CGGTACCGTTGTAC 815

RESULT 11
 ECOUBIA
 LOCUS
 DEFINITION 1498 bp DNA linear BCT 28-MAR-1994
 gene, 5' end and chorismate lyase (ubiC) gene, complete cds.
 ACCESSION M96268
 VERSION M96268.1 GI:347886
 KEYWORDS 4-hydroxybenzoate octaprenyl transferase; aerobic respiratory deficiency; chorismate lyase; membrane-bound protein; prenyltransferase; transferase; ubiA gene; ubiC gene; ubiquinone biosynthesis.

SOURCE
 ORGANISM Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 1498)
 AUTHORS Wu, G., Williams, H. D., Gibson, F. and Poole, R. K.
 TITLE Mutants of Escherichia coli affected in respiration: the cloning and nucleotide sequence of ubiA, encoding the membrane-bound p-hydroxybenzoate:octaprenyltransferase
 JOURNAL J. Gen. Microbiol. 139 (Pt 8), 1795-1805 (1993)
 PUBMED 8409922
 FEATURES Location/Qualifiers
 source 1..1498
 /organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /strain="K-12"
 /db_xref="taxon:562"
 /map="91.8 min"
 /tissue_lib="cosmid"
 <1..524
 /note="ORF; the sequence before nucleotide number 110 might not be genuine. Three Clai sites, which is the enzyme used for cloning, were identified in this region, so it could be a cloning artifact"
 /codon_start=3
 /transl_table=11
 /protein_id="AAAI7026.1"
 /db_xref="GI:347887"
 /translation="RFSEVNFDTQQMIDFINFNTKNKWLIDKNNFFTLIDQALRSTDMTKANVHLKIEWIRSDVSPIFETEDNLRFTNELTRNDNIFLPSVDDGPVMTVSSQRLHDLNAPTQDTNNSTVIYKSRHEMLPVNLTQETLFSSKSHGKVALPIFTASWRAHRIMKGV"
 RBS 735..738
 gene 747..1244
 /gene="ubiC"
 CDS 747..1244
 /gene="ubiC"
 /codon_start=1
 /transl_table=11
 /product="chorismate lyase"
 /protein_id="AAAI7027.1"
 /db_xref="GI:347888"
 /translation="MSPALTLQALRYCKEIPALDPQLLDMLLEDSTMTRFESQOGKTVSVTMIREGFVQNEIPEDLPKESRYLWREILLICADGEPWLACRTVVPVSTLSGPELALQKLKLTPLGLRYLFTSTLTRDFIEGRDAGLWGRRLRLSGKPLLTLELFLPASPLY"
 gene 1257..1498
 /gene="ubiA"
 CDS 1257..1498
 /gene="ubiA"
 /note="The last part of this ORF is to be found in entry K00127.; ORF"
 /codon_start=1
 /transl_table=11
 /product="4-hydroxybenzoate-octaprenyl transferase"

/protein_id="AA17028.1"
/db_xref="GI:347889"
/translation="MEWSLTQNKLLAFHRLMRDKPIGALLLLWPTLWALWATPGVP
QLWILAVFVAGVWLMFAAGCVVNDYADRFDFGHVKET"

ORIGIN

Alignment Scores:
Pred. No.: 8,59e-80 Length: 1498
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 15 Gaps: 0

US-10-718-311-16 (1-170) x ECOUBIA (1-1498)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
DB 747 ATGTACACCCCGGTTAAGCAACTGCGTGGCTCGCTATTGTAAGAGATCCCTGCC 806
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
DB 807 CTGGATCCGCAACTGCTCACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTGA 866
QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
DB 867 CAGCAGGGGAAAAACGGTAAGCGTGACATGATCCGCGAAGGTTTGTCCAGCAGATGAA 926
QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
DB 927 ATCCCGGAAGAACTGCGCGTGTGCGGAAGAGTCTCGTTACTGGTACGTAATTTTG 986
QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
DB 987 TTATGTGCGCGATGGTGAACCGTGGCTTGGCGGTGCTGCTGCTGTCAAGTTA 1046
QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
DB 1047 AGCGGCGCGAGTGGCGTTACAAAATTTGGGTAAACCGCTTAGGACGCTATCTGTT 1106
QY 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
DB 1107 ACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGCGTGATGCGCGGTG 1166
QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
DB 1167 CGAGCTTCCCGCTCGGATTAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTAC 1226
QY 166 AlaSerProLeuTyr 170
DB 1227 GCGTCACCGTTGTAC 1241

RESULT 12

E11274
LOCUS ubiC and ubiA gene.
DEFINITION E11274
ACCESSION E11274
VERSION E11274.1 GI:22024916
KEYWORDS JP 1996107789-A/1.
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
REFERENCE 1 (bases 1 to 2000)
AUTHORS Matsuda,H., Kawamuki,M. and Nakagawa,T.
TITLE PRODUCTION OF UBIQUINONE-10
JOURNAL Patent: JP 1996107789-A 1 30-APR-1996;
COMMENT OS Escherichia coli
PN JP 1996107789-A/1
PD 30-APR-1996
PF 13-OCT-1994 JP 1994273071
PI MATSUDA HIDEYUKI, KAWAMUKI MAKOTO, NAKAGAWA TSUYOSHI PC

C12N15/09, A61K31/12, A61K31/12, C07H21/04, C12N1/21, C12P7/66, PC
(C12N1/21,
PC C12R1.01), (C12P7/66, C12R1.01);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH source 1..2000
FT /organism='Escherichia coli'
FT CDS 380..877
FT /product='UbiC protein'
FT CDS 890..1762
FT /product='UbiA protein'.
FEATURES
source
1..2000
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"

ORIGIN

Alignment Scores:
Pred. No.: 1,17e-79 Length: 2000
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x E11274 (1-2000)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
DB 380 ATGTACACCCCGGTTAAGCAACTGCGTGGCTCGCTATTGTAAGAGATCCCTGCC 439
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
DB 440 CTGGATCCGCAACTGCTCACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTGA 499
QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
DB 500 CAGCAGGGGAAAAACGGTAAGCGTGACATGATCCGGAAGGTTTGTCCAGCAGATGAA 559
QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
DB 560 ATCCCGGAAGAACTGCGCGTGTGCGGAAGAGTCTCGTTACTGGTACGTAATTTTG 619
QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
DB 620 TTATGTGCGGATGGTGAACCGGCTTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 679
QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
DB 680 AGCGGCGCGAGTGGCGTTACAAAATTTGGGTAAACCGCTTAGGACGCTATCTGTT 739
QY 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
DB 740 ACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGCGTGATGCGCGGTG 799
QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
DB 800 CGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTAC 859
QY 166 AlaSerProLeuTyr 170
DB 860 GCGTCACCGTTGTAC 874

RESULT 13

ECOUBICA 2000 bp DNA linear BCT 26-APR-1993
LOCUS Escherichia coli chorismate lyase (ubiC), 4-hydroxybenzoate
DEFINITION octaprenyl transferase (ubiA) genes, complete cds, and
sn-glycerol-3-phosphate acyltransferase (plsB) genes, 3' end.
ACCESSION M93136

AUTHORS Lightner,V.A., Bell,R.M. and Modrich,P.
 TITLE The DNA sequences encoding plsB and dgk loci of Escherichia coli
 JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)
 PUBMED 6309817
 REFERENCE 2 (bases 1 to 2000)
 AUTHORS Nichols,B.P. and Green,J.M.
 TITLE Cloning and sequencing of Escherichia coli ubiC and purification of chorismate lyase
 JOURNAL J. Bacteriol. 174 (16), 5309-5316 (1992)
 PUBMED 1644758
 COMMENT Original source text: Escherichia coli (sub_strain W3110, strain X-12) (library: Kohara lambda) DNA.
 FEATURES
 source
 1..2000
 /organism="Escherichia coli W3110"
 /mol_type="genomic DNA"
 /strain="K-12"
 /sub_strain="W3110"
 /db_xref="taxon:316407"
 /map="91.8 min"
 /tissue_lib="Kohara lambda"
 380..877
 /gene="ubiC"
 380..877
 /gene="ubiC"
 /function="enzymatic chorismate --> p-hydroxybenzoate + pyruvate"
 /experiment="experimental evidence, no additional details recorded"
 /note="N-terminal amino acid sequence of chorismate lyase :NH2-SHPALQLRALRYCEIPAL-"
 /citation={2}
 /codon_start=1
 /transl_table=11
 /product="chorismate lyase"
 /protein_id="AAA24716.1"
 /db_xref="GI:148107"
 /translation="MSHPALQLRALRYCKEIPALDQPOLLDWLLLESDMTRKFEQQKG TVSTMIRGFEVNEIPEELPLPKESRYWLEILLCADGSPWLAGRTVVPVSTLSG PELALQKLGKPLGRYLTSSLTITDFIEIGRDAGLWGRSRLRLSGRPLLLITELFLP ASPLY"
 890..1762
 /gene="ubiA"
 890..1762
 /gene="ubiA"
 /function="enzymatic - 3-octaprenyl-4-hydroxybenzoate synthesis"
 /experiment="experimental evidence, no additional details recorded"
 /citation={2}
 /number=2
 /codon_start=1
 /transl_table=11
 /product="4-hydroxybenzoate-octaprenyl transferase"
 /protein_id="AAA24717.1"
 /db_xref="GI:148108"
 /translation="MEWSLTQNKLLAFHRLMETDKPIGALLLLWPTLWLVATPGVP QLWILAVFAGVWLMRAGCVVNDYADRFQGHVVRVTANRPLPSGAVTEKRALFVV LVLSFLVLTUNTLISLAALAWYIPMKRYTHLPQVVLGAAGFWSIPMAFAA VSESFLSCWLMFLANLWAVDQYAVMDRDDDKIKGISTAILFGYDKLIIGIL QIGVLAMAIIGELNGLGWYWSILVAGALFVYQOKLIANRERACFKAFMNNYVG LVFLGLAMSYWHF"
 complement(1917..2000)
 /gene="plsB"
 complement(1917..2000)
 /gene="plsB"
 /EC_number="2.3.1.15"
 /function="lysophosphatidic acid synthesis"
 /experiment="experimental evidence, no additional details recorded"
 /citation={1}
 /codon_start=1
 /transl_table=11

/product="sn-glycerol-3-phosphate acyltransferase"
 /protein_id="AAA24718.1"
 /db_xref="GI:148109"
 /translation="MKVYQLLAELITSVDRLTIESATQGEG"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,17e-79 Length: 2000
 Score: 854.00 Matches: 165
 Percent Similarity: 100.0% Conservatives: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 96.3% Indels: 0
 DB: 15 Gaps: 0
 US-10-718-311-16 (1-170) x ECUUBIPLS (1-2000)
 QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
 |||||
 Db 380 ATGTACACCCCGCGTTAACGCAACTGCGTGGCGCTATTGTAAAGAGATCCCTGCC 439
 |||||
 QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
 |||||
 Db 440 CTGGATCCGCAACTGCTGACTGGCTGTTCGGAGGATTCATGACAAAACGTTTGA 499
 |||||
 QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
 |||||
 Db 500 CAGCAGGGGAAAACCGTAAGCGTGACGATGATCCGGAAGGGTTGTCGAGCAGATGAA 559
 |||||
 QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 85
 |||||
 Db 560 ATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 619
 |||||
 QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
 |||||
 Db 620 TTATGTGCGATGGTGAACCGTGGCTTCCCGTGTACCGTCTCTGTGCAACGTTA 679
 |||||
 QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
 |||||
 Db 680 AGCGGGCGGAGCTCGCGTTACAAAATTGGGTAAACCGCCCTAGGACGCTATCTGTTTC 739
 |||||
 QY 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 145
 |||||
 Db 740 ACATCATCGACATTAAACCGGACTTTATTGATAGTAGCCGTGATGCCGGCTGTGGGG 799
 |||||
 QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
 |||||
 Db 800 CGACGTTTCCCGCTCGATTAAACCGGTAACCGCTGTTGCTAACAGAACTGTTTTTACC 859
 |||||
 QY 166 AlaSerProLeuTyr 170
 |||||
 Db 860 GCGTCACCGTTGTAC 874
 |||||
 RESULT 15
 ECUBIAC
 LOCUS
 DEFINITION E.coli ubiC and ubiA genes for chorismate lyase and 4-hydroxybenzoate octaprenyltransferase.
 ACCESSION X57434
 VERSION X57434.1 GI:43233
 KEYWORDS 4-hydroxybenzoate octaprenyl transferase; 4-hydroxybenzoate synthetase; chorismate lyase.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCE 1
 AUTHORS Nishimura,K., Nakahigashi,K. and Inokuchi,H.
 TITLE Location of the ubiA gene on the physical map of Escherichia coli
 JOURNAL J. Bacteriol. 174 (17), 5762 (1992)
 PUBMED 1512213
 REFERENCE 2 (bases 1 to 2034)
 AUTHORS Nishimura,K.
 TITLE Direct Submission

JOURNAL Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto University, Sakyo-ku, Kyoto 606, Japan

FEATURES Location/Qualifiers

source 1..2034

/organism="Escherichia coli"

/mol_type="genomic DNA"

/strain="K12 W3110"

/db_xref="taxon:562"

/map="91.7 min"

/clone="634"

/clone_lib="Kohara library"

137..142

-35_signal

159..164

-10_signal

291..899

gene /gene="ubiC"

291..899

CDS /gene="ubiC"

/codon_start=1

/transl_table=11

/product="4-hydroxybenzoate synthetase"

/protein_id="CAA40681.1"

/db_xref="GI:43234"

/db_xref="GOA:P26602"

/translation="MRLRFCCVLDDHLCFTSPVNTFLRYNAFTLCNCGFQMSHPALT

OLRLRYCKEIPALDPOILLLEDSDMTKEFEOGKTVSVTMIKEGFVEQNEIPEEL

RLLPKESRYWREILLCADGEPLWLAGRVVPVSTLSGPFLALQKLKLTPLGRYLFTSS

TLTRDFEIGRDAGLWGRSRRLUSGKPLLLTEULFLPASPLY"

912..1784

gene /gene="ubiA"

912..1784

CDS /gene="ubiA"

/codon_start=1

/transl_table=11

/product="4-hydroxybenzoate octaprenyltransferase"

/protein_id="CAA40682.1"

/db_xref="GI:43235"

/db_xref="GOA:P26601"

/translation="MWSLTQNKLAFLHRLMRTDKPIGALLLLMPTLWALMVATPGVP

QLWILAVFVAGVLMRAAGCVVNDYADKFDGHVKRTANRPLPSGAVTEKEARALFVV

LVLISFLVLTNTMTILLSIAALAWVPFMKRYTHLPQVILGAAFGWSIPMAFAA

VSESPLSCWLMFLANILWAVYDQYAMVDRDDVKIGIKSTAILFGYDKLIIGIL

QIGVLALMIIIGELNGLGWYNSILVAGALFVYQQKLIANRERECFKAPMNNNYVG

LVLFLGLANSIWHF"

ORIGIN

Alignment Scores:

Pred. No.:	1.2e-79	Length:	2034
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.3%	Indels:	0
DB:	15	Gaps:	0

US-10-718-311-16 (1-170) x ECUBIAC (1-2034)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25

Db 402 ATGTACACCCCGCGTTAAACGCACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 461

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45

Db 462 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGA 521

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnGlnAsnGlu 65

Db 522 CAGCAGGGGAAAAACGGTAAGCGGTGACGATGATCCCGAAGGGTTTGTTCGACAGAAATGAA 581

Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85

Db 582 ATCCCGGAAGAACTGCCGCTGTCGCCGAAAGAGTCTCGTTACTGCTTACGTGAAATTTG 641

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105

Db 642 TTAITGTCGGATGGTGAAACCGTGGCTTGCCTGCTACCGTCTCTGTGTCAACGTTA 701

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125

Db 702 AGCGGGCGGAGCTGGCGTTACAAAAAATTGGGTAAAAACGCCGTTAGGACGCTATCTGTTT 761

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145

Db 762 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 821

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 165

Db 822 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTGTCTTAACAGAACTGTTTTTACCG 881

Qy 166 AlaSerProLeuTyr 170

Db 882 GCGTCACCGTTGTAC 896

Search completed: May 30, 2006, 05:52:57

Job time : 5159.27 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2006, 11:01:13 ; Search time 207.045 Seconds
(without alignments)
2304.491 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVWMSHPALTQLRALRYC.....RLSGKPLLTLELFPASPLY 170

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/abs8/ABSSWEB/spool/US10718311/runat_26052006_164922_13534/app_query.fasta_1
-DB=issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs02p -USER=US10718311 @CGN 1.1 615 @runat_26052006_164922_13534
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents NA:*
1: /EMC_Celerra_IDS3/ptodata/2/ina/1 COMB.seq:*
2: /EMC_Celerra_IDS3/ptodata/2/ina/5 COMB.seq:*
3: /EMC_Celerra_IDS3/ptodata/2/ina/6A COMB.seq:*
4: /EMC_Celerra_IDS3/ptodata/2/ina/6B COMB.seq:*
5: /EMC_Celerra_IDS3/ptodata/2/ina/7 COMB.seq:*
6: /EMC_Celerra_IDS3/ptodata/2/ina/H COMB.seq:*
7: /EMC_Celerra_IDS3/ptodata/2/ina/PCUTUS COMB.seq:*
8: /EMC_Celerra_IDS3/ptodata/2/ina/PP COMB.seq:*
9: /EMC_Celerra_IDS3/ptodata/2/ina/RE COMB.seq:*
10: /EMC_Celerra_IDS3/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	887	100.0	684	3	US-09-896-866B-16
2	887	100.0	684	3	Sequence 16, Appl
3	854	96.3	495	3	Sequence 7, Appl
4	854	96.3	495	3	Sequence 12, Appl
5	854	96.3	854	3	Sequence 3, Appl
6	854	96.3	6841	3	Sequence 24, Appl
7	673	75.9	645	3	Sequence 25, Appl
8	401	45.2	537	3	Sequence 1234, Ap
					Sequence 3122, Ap

9	186	21.0	630	3	US-09-252-991A-7442	Sequence 7442, Ap
10	174.5	19.7	519	3	US-09-540-236-1889	Sequence 1889, Ap
11	171.5	19.3	269223	3	US-09-596-002-41	Sequence 41, Appl
12	159	17.9	540	3	US-09-328-352-343	Sequence 343, Appl
13	146.5	16.5	1248	3	US-09-252-991A-7361	Sequence 7361, Ap
c 14	121	13.6	195	3	US-09-543-681A-3086	Sequence 3086, Ap
15	90	10.1	2127	3	US-09-818-780-75	Sequence 75, Appl
16	85	9.6	888	3	US-09-107-433-1721	Sequence 1721, Ap
c 17	83	9.4	84839	3	US-09-949-016-15816	Sequence 15816, A
18	82	9.2	601	3	US-09-949-016-68693	Sequence 68693, A
19	82	9.2	807	3	US-09-583-110-1232	Sequence 1232, Ap
20	81	9.1	1176	3	US-09-252-991A-10640	Sequence 10640, A
21	81	9.1	1581	3	US-09-252-991A-10731	Sequence 10731, A
c 22	81	9.1	2055	3	US-09-252-991A-10433	Sequence 10433, A
c 23	81	9.1	62354	3	US-09-949-016-16188	Sequence 16188, A
c 24	80	9.0	6463	2	US-08-962-284-3	Sequence 3, Appl
c 25	80	9.0	68778	3	US-09-949-016-16406	Sequence 16406, A
c 26	80	9.0	72455	3	US-09-949-016-13793	Sequence 13793, A
c 27	80	9.0	77535	3	US-09-949-016-14279	Sequence 14279, A
c 28	80	9.0	77535	3	US-09-949-016-14280	Sequence 14280, A
c 29	80	9.0	77535	3	US-09-949-016-14281	Sequence 14281, A
c 30	80	9.0	104475	3	US-09-949-016-12115	Sequence 12115, A
c 31	80	9.0	111282	3	US-09-754-250-3	Sequence 3, Appl
c 32	80	9.0	111282	3	US-10-094-989-3	Sequence 3, Appl
c 33	79.5	9.0	47981	3	US-09-679-279-1	Sequence 1, Appl
c 34	79.5	9.0	4403765	3	US-09-103-840A-2	Sequence 2, Appl
c 35	79.5	9.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
c 36	79	8.9	5246	3	US-09-799-451-474	Sequence 474, Appl
37	78.5	8.9	1486	2	US-08-250-859-1	Sequence 1, Appl
38	78.5	8.9	1486	2	US-08-347-029-1	Sequence 1, Appl
39	78.5	8.9	1486	2	US-08-490-803-1	Sequence 1, Appl
40	78.5	8.9	1486	2	US-08-457-254-1	Sequence 1, Appl
41	78.5	8.9	1486	2	US-08-484-257-1	Sequence 1, Appl
42	78.5	8.9	1486	3	US-08-999-927-3	Sequence 3, Appl
43	78.5	8.9	1486	3	US-08-461-819-3	Sequence 1, Appl
44	78.5	8.9	1486	7	PCT-US94-08806-1	Sequence 1, Appl
45	78.5	8.9	1486	7	PCT-US95-01775-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-896-866B-16
; Sequence 16, Application US/09896866B
; Patent No. 6642036
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein
US-09-896-866B-16

Alignment Scores:
Pred. No.: 1.94e-100 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0

DB: 3 Gaps: 0
US-10-718-311-16 (1-170) x US-09-896-866B-16 (1-684)
Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Dy 172 ATGCAGGTGGCATATGTACACCCCGGTTAACGCAACTGCGTGGCTCGCTATTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Dy 172 ATGCAGGTGGCATATGTACACCCCGGTTAACGCAACTGCGTGGCTCGCTATTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Dy 232 AAAGAGATCCCTGCCCTGCACTGCTCGACTGGCTGGCTGGAGGATTCATG 291
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Dy 292 ACAAAACGTTTGAACACAGCAGGAGGAAAAACGGTAAGCGTGAACGATCCGCGAAGGGTTT 351
Qy 61 ValGluGlnAnGluileProGluLeuProLysProLysGluSerArgTyrTrp 80
Dy 352 GTCGAGCAGATGAATCCCGGAAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Dy 412 TTACGTGAATTTTGTATGTATGCGCATGCTGAACCGTGGCTTGGCGGTCTGCTCGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Dy 472 CCTGTGTCAACGTTAAGCGGGCGGAGCTGGCGTTACAAAAATTGGGTAAAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerThrLeuThrArgAspPheIleGluileGlyArgAsp 140
Dy 532 GGACGCTATCTGTTACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Dy 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTCGGATTAAGCGGTAAACCGCTGTGTGTAACA 651
Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Dy 652 GAACGTGTTTACCGGGCTCACCGTTGTAC 681

RESULT 2
US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. 6683231
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUZ
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7

Alignment Scores:
Pred. No.: 1.94e-100 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-855-341-7 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Dy 172 ATGCAGGTGGCATATGTACACCCCGGTTAACGCAACTGCGTGGCTCGCTATTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Dy 232 AAAGAGATCCCTGCCCTGCACTGCTCGACTGGCTGGCTGGAGGATTCATG 291
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Dy 292 ACAAAACGTTTGAACACAGCAGGAGGAAAAACGGTAAGCGTGAACGATCCGCGAAGGGTTT 351
Qy 61 ValGluGlnAnGluileProGluLeuProLysProLysGluSerArgTyrTrp 80
Dy 352 GTCGAGCAGATGAATCCCGGAAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Dy 412 TTACGTGAATTTTGTATGTATGCGCATGCTGAACCGTGGCTTGGCGGTCTGCTCGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Dy 472 CCTGTGTCAACGTTAAGCGGGCGGAGCTGGCGTTACAAAAATTGGGTAAAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerThrLeuThrArgAspPheIleGluileGlyArgAsp 140
Dy 532 GGACGCTATCTGTTACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Dy 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTCGGATTAAGCGGTAAACCGCTGTGTGTAACA 651
Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Dy 652 GAACGTGTTTACCGGGCTCACCGTTGTAC 681

RESULT 3
US-09-896-866B-12
; Sequence 12, Application US/09896866B
; Patent No. 6642036
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; TITLE OF INVENTION: Benzoic Acid Glucosides
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12

Alignment Scores:
Pred. No.: 1.48e-96 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-896-866B-12 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 25
Dy 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTATTGTTAAAGAGATCCCTGCC 60

Qy 26 LeuAspProGluLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTTGAA 120
Qy 46 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 65
Db 121 CAGCAGGAGAAACGGTAAGCGTGACGATATCCGCGAAGGGTTCGTCAGCAGAAATGAA 180
Qy 66 IleProGluGluLeuProLeuProLysGluSerArgTyTrpLeuArgGluIleLeu 85
Db 181 ATCCCGGAGAAAGTCCGCTGCTCCGGAAGAGTCTCGTTACTGGTTACGTGAATTTTG 240
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTGCGGATGTAACCGTGGCTTGGCGGTGCTACCGTCTGCTGTCAACGTTA 300
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyTrpLeuPhe 125
Db 301 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAAACCGGACTTTATTTAGATAGGCGGTGATCCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGACGTTCCCGCTGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 480
Qy 166 AlaSerProLeuTytr 170
Db 481 GCGTCACCGTTGTAC 495

RESULT 4
US-09-855-341-3
; Sequence 3, Application US/09855341
; Patent No. 6683231
; GENERAL INFORMATION:
; APPLICANT: MEYER, KNUIT
; APPLICANT: MEYER, KNUIT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZAZOIC ACID
; FILE REFERENCE: IN GREEN PLANTS
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-09-855-341-3

Alignment Scores:
Pred. No.: 1.48e-96 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-855-341-3 (1-495)
Qy 6 MetSerHieProAlaLeuThrGlnLeuArgAlaLeuArgTyCysLysGluLeuProAla 25
Db 1 ATGTACACCCCGGTAAACGTAACGTCGCTGCGCTATTTGAAAGAGATCCCTGCC 60
Qy 26 LeuAspProGluLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTTGAA 120

Qy 46 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 65
Db 121 CAGCAGGAGAAACGGTAAGCGTGACGATATCCGCGAAGGGTTCGTCAGCAGAAATGAA 180
Qy 66 IleProGluGluLeuProLeuProLysGluSerArgTyTrpLeuArgGluIleLeu 85
Db 181 ATCCCGGAGAAAGTCCGCTGCTCCGGAAGAGTCTCGTTACTGGTTACGTGAATTTTG 240
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTGCGGATGTAACCGTGGCTTGGCGGTGCTACCGTCTGCTGTCAACGTTA 300
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyTrpLeuPhe 125
Db 301 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAAACCGGACTTTATTTAGATAGGCGGTGATCCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGACGTTCCCGCTGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 480
Qy 166 AlaSerProLeuTytr 170
Db 481 GCGTCACCGTTGTAC 495

RESULT 5
US-09-064-693A-24
; Sequence 24, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSER: W. Gary Goodson
; ADDRESSES: INEL--Lockheed Martin Idaho
; ADDRESSES: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-064-693A-24

```
Alignment Scores:
Pred. No.: 3,39e-96 Length: 854
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-064-693A-24 (1-854)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 321 ATGTCACACCCCGGTTAACGCACTGCTGCGCTGCTATTGTAAGAGATCCCTGCC 380
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 440
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 441 CAGCAGGGGAAAACCGTTAAGCGTGACGATGATCCGCGAAGGTTTGTCTCAGCAGAAATGA 500
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 501 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGCTTACGTGAAATTTTG 560
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrIleu 105
Db 561 TTATGTGCGGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTGTTCTGTCAACGTTA 620
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuphe 125
Db 621 AGCGGGCGGAGCTGGCGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 680
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 681 ACATCATCGACATTAACCCGGAGCTTTATTGAGATAGCCGCTGATGCCGCTGTGGGG 740
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 741 CGAGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800
Qy 166 AlaSerProLeuTyr 170
Db 801 CGGTCACCGTTGTAC 815

RESULT 6
US-09-064-693A-25
; Sequence 25, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-064-693A-25

Alignment Scores:
Pred. No.: 7,76e-95 Length: 6641
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-064-693A-25 (1-6641)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 321 ATGTCACACCCCGGTTAACGCACTGCTGCGCTGCTATTGTAAGAGATCCCTGCC 380
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 440
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 441 CAGCAGGGGAAAACCGTTAAGCGTGACGATGATCCGCGAAGGTTTGTCTCAGCAGAAATGA 500
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 501 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGCTTACGTGAAATTTTG 560
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrIleu 105
Db 561 TTATGTGCGGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTGTTCTGTCAACGTTA 620
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuphe 125
Db 621 AGCGGGCGGAGCTGGCGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 680
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 681 ACATCATCGACATTAACCCGGAGCTTTATTGAGATAGCCGCTGATGCCGCTGTGGGG 740
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 741 CGAGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800
Qy 166 AlaSerProLeuTyr 170
Db 801 CGGTCACCGTTGTAC 815

RESULT 7
US-09-489-039A-1234
; Sequence 1234, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
```

Alignment Scores:	
Pred. No.:	2,74e-40
Score:	401.00
Percent Similarity:	68.6%
Best Local Similarity:	57.9%
Query Match:	45.2%
DB:	3
Length:	537
Matches:	81
Conservative:	15
Mismatches:	44
Indels:	0
Gaps:	0

US-10-718-311-16 (1-170) x US-09-543-681A-3122 (1-537)

Qy 31 LeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThr 50
 ||| ||| ||| |||
 ::: ::: ::: :::
Db 109 TTAAGTTGGCTGCTGGAATTAGTTCAAACACACGCGTTTTTGAGCAACATTGCATCAA 168

QY 51 valser val Thr Met Ile Arg Gly Tyr phe val Glu Glu Asn Glu Ile pro Glu Glu Leu 70

[illegible]

Db 229 AAGTGTTCCTTATAGCCAGCGTATTGGCTAGAGAAATCGTCTTTGTGGGATAAT 288

Q7 289 GTACCTTGTTATTAGGCGGAACATTAGTCCGGAAGACATTACGGGTGAAGATCGC 348
Db

Qy 111 AlaLeuGlnIysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerSerThrLeu 130

Db 349 CAATTAGTGAATTTGGGTACGGTGCATTAGGACGTTAICTGTTCACAGGAACAACCTTA 408

QY 131 *inr*ATGAspnnellresllulleGlyARGspAlaglyLeuIRpGlyARGargseARGLeu 150
:::|||||:::
Db 409 AGTCGTGATTTTATTCATATTTGGGCAACAATAATGGACATTTGGTTACGTCTTCCGTTTC 468

[illegible]

D_b 469 CAGCTTTCAGATAAACCTTTATTATTAACTGAGGTGTTTTGGCTGCGCATCACCTGTAT 528

RESULT 9
US-09-252-991A-7442
; Sequence 7442, Application US/09252991A

Patent No. 6551795
GENERAL INFORMATION:
PATENT CLASS. No. 7
INT. CLASS. No. 7
PUBL. NO. 6551795

APPELLANI: MARC J. RUBENFELD et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/352,991A
;
; CURRENT FILING DATE: 1999-02-18

```

PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; NUMBER OF SEQ ID NOS: 33142

```

; SEQ ID NO 7442
;
; LENGTH: 630
;
; TVDP. DNA

```

ORGANISM: *Pseudomonas aeruginosa*
US-09-252-991A-7442

Alignment Scores:
Pred. No.: 1.67e-13 Length: 630

Score:	186.00	Matches:	50
Percent Similarity:	43.0%	Conservative:	21
Percent Similarity:	20.2%	Midrange:	54
Percent Similarity:	20.2%	Liberal:	54

```

Best Local Similarity: 30.3%
Query Match: 21.0%
DB: 3
Indels: 30
Gaps: 3
Mismatches: 64

```

```
US-10-718-311-16 (1-170) x US-09-252-991A-7442 (1-630)
Qy 22 GluileProAlaLeuAspProGlnLeuLeuAspTrrpLeuLeuLeuGluAspSerMetThr 41
Db 127 CAACGTGCAAGGCTCCCGCGCGTCTCGACGTGGCTGTCGACGAGGGTCTGCTGACC 186
Qy 42 LysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVal 61
Db 187 CGGGGC-----CTGACCGCCTGGCGCGCGCGGCTTC 219
Qy 62 GluGlnAenGluileProGluLeuLeuProLeuLeu----- 73
Db 220 CGCGTCGAACCCCTGCTGAAGGTGGCGAGACCTGCGCGCAGCAAGATCCAGGGGGCTC 279
Qy 74 -----ProLysGluSerArgTyrTrrpLeuArgGluileLeuLeuCysAlaAspGlyGlu 91
Db 280 GACGTCCCTACCGCAGTAGCGGTGGGTGGCGGAGGTCACCTGCATGGCCACGACCGT 339
Qy 92 ProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAla 111
Db 340 CCCTGGGTGTCGCGCGCAGGTGGCGCGCGCAGCGCCCTCGAAGGCTCGGGCTTCGAC 399
Qy 112 LeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerThrLeuThr 131
Db 400 CTGGCGTGTCTGGCAGCCCGCTCGCTGGCGGAGTGTGTTTCAGCAGCAGCGCTTCGAG 459
Qy 132 ArgAspPheIleGluile-----GlyArgAsp 140
Db 460 CGCGGGCCATCGAAGTCTCGGTATCCGGCGCGGTCTGCCCGCGAGGTCCGGCC 519
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 520 GAGGCTCTCTGGCGCGTCTCGTCAAGTTCCTCCCGCGCGCGCTCGGGTCTGTTGGCG 579
Qy 161 GluLeuPheLeuPro 165
Db 580 GAGGTGTACCTACCG 594

RESULT 10
US-09-540-236-1889
; Sequence 1889, Application US/09540236
; Patent No. 667310
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709-2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1889
; LENGTH: 519
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-1889

Alignment Scores:
Pred. No.: 3,32e-12 Length: 519
Score: 174.50 Matches: 48
Percent Similarity: 49.4% Conservative: 36
Best Local Similarity: 28.2% Mismatches: 57
Query Match: 19.7% Indels: 29
DB: 3 Gaps: 7

US-10-718-311-16 (1-170) x US-09-540-236-1889 (1-519)
Qy 20 CysLysGluileProAlaLeuAsp----- 27
Db 4 TGTAACATTGTACCGCCCAATAGATGATTTTCATGCCAATGGATGATATTAGACCAC 63
Qy 28 -----ProGlnLeuLeuAspTrrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 280 -----ProGlnLeuLeuAspTrrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
```

```
Db 64 CTGCCTCAAAACCTTGCCGATTTTGTATTACTCAAGGTCGTTAAACGGCAGGATTAGAA 123
Qy 46 GlnGlnGly---LysThrValSerValThrMetIleArgGluGly----- 59
Db 124 GCACACTCAAAACCCCTAAAGGTTTGAATCATCATCAAGGCTATCAGCCACTCACC 183
Qy 60 PheValGluGlnAenGluileProGluLeuProLeuLeuProLysGluSerArgTyr 79
Db 184 TTTGCCCAAAATCACAACCTG-----GGACTGCC-----CTAAATCGCCCAAAATGGCT 234
Qy 80 TrpLeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrVal 99
Db 235 TGGTAGCGAGTGTCAAGTTATACGGTATCGGACGCTTGGGTGCTTGCCTCCAAAGTATT 294
Qy 100 ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro 119
Db 295 TTTCTTTTGTTCATCACTGACAGGTTTCATTAGCAGCTTAAAGCAGCTGGGTGCAACGCC 354
Qy 120 LeuGlyArgTyrLeuPheThrSerSer-----ThrLeuThrArgAspPheIleGlu 136
Db 355 ATCGGATATGTTTGTTCAAAAAAGGCAACACTGCCAATTAAAGCGTCATTTTAT--- 411
Qy 137 IleGlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysPro 156
Db 412 -----CGGTATGATGTCCTCAATTTGGTGGCAACAGTTTATGATATCAATGGCGGTAAA 465
Qy 157 LeuLeuLeuThrGluLeuPheLeuProAla 166
Db 466 ATCTTGATGATGAGCTATTTTGGCAGCT 495
```

RESULT 11

```
US-09-596-002-41
; Sequence 41, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 269223
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 41
; PUBLICATION INFORMATION:
US-09-596-002-41
```

```
Alignment Scores:
Pred. No.: 1.09e-07 Length: 269223
Score: 171.50 Matches: 46
Percent Similarity: 53.3% Conservative: 34
Best Local Similarity: 30.7% Mismatches: 55
Query Match: 19.3% Indels: 15
DB: 3 Gaps: 6
```

US-10-718-311-16 (1-170) x US-09-596-002-41 (1-269223)

```
Qy 26 LeuAspProGlnLeuLeuAspTrrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 16824 CTGCCTCAAAACCTTGCCGATTTTGTATTACTCAAGGTCGTTAAACGGCAGGATTAGAA 16883
Qy 46 GlnGlnGly---LysThrValSerValThrMetIleArgGluGly----- 59
Db 16884 GCACACTCAAAACCCCTAAAGGTTTGAATCATCATCAAGGCTATCAGCCACTCACC 16943
```

Qy	100	ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro	119
Db	298	TTTCCAATTCCAAAGTTTACAAAAAAGCCGCTATATTTTCAGCATATGGTCTCTAAGCCG	357
Qy	120	LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArg	139
Db	358	ATAGTCTTTTTTATTTCAAAGAACAACACCACTTTGTGATCGCCGGGTATTTCGTTTA	417
Qy	140	AspAlaGlyLeuTyrGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeu	159
Db	418	CCAGAAGGC---TGGACCGCAAAAGTTGTTATATCTTGCATGGATGTAAATTTATTGTT	474
Qy	160	ThrGluLeuPheLeuProAla	166
Db	475	CAAGAAACATTCTTACCGGCT	495

RESULT 13
US-09-252-991A-7361
; Sequence 7361, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

: FILE REFERENCE: 10/196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 7361
: LENGTH: 1248
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7361

```

```

Alignment Scores:
Pred. No.:          3,78e-08          Length:          1248
Score:             146.50           Matches:          35
Percent Similarity: 46.4%           Conservative:    10
Best Local Similarity: 36.1%        Mismatches:      41
Query Match:       16.5%           Indels:          11
DB:                3                Gaps:            1

US-10-718-311-16 (1-170) x US-09-253-991A-7361 (1-1248)

```

```

Qy      80  TrpLeuArgGluIleLeuLeuCysAlaAspGlyGluProTrrpLeuAlaGlyArgThrVal 99
      |||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::
Db      5  TGGGTCCGCGAGGTCCTACCTGCATGGCCACGACCGTCCTGGGTCTTCCGCCGACGGTG 64

Qy     100  ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro 119
      |||:::||||:::||||:::||||:::||||:::||||:::||||:::
Db     65  GCGCGCGCGACGCGCCTGGAAGGCTCGGCGCTTCGACCTGGCGGCTCTCGGCACCGCGTCG 124

Qy     120  LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIle----- 137
      |||:::||||:::||||:::||||:::||||:::||||:::||||:::
Db     125  CTGGCGGAGTTGCTGTTTCAGGCAGCGCCTTCGAGCGCGGCGCCCATCGAAGCTGCGCGC 184

Qy     138  -----GlyArgAspAlaGlyLeuTrpGlyArgArgSer 148
      |||:::||||:::||||:::||||:::||||:::||||:::||||:::
Db     185  TATCCGCGCGCGGTCTGCCCGCGAGGTCCGCGCGGAGGTCTCTGGGGCGTCTCGCTCA 244

Qy     149  ArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
      |||:::||||:::||||:::||||:::||||:::||||:::||||:::
Db     245  CGGTTTTCGCGCGCGCTCGGGTCTGTGTGGCGAGGTGTACCTACCG 295

RESULT 14
US-09-543-681A-3086/c
; Sequence 3086, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:

```

```

; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001

```

Alignment Scores:	3.25e-06	Length:	195
Pred. No.:	121.00	Matches:	24
Score:	82.4%	Conservative:	4
Percent Similarity:	70.6%	Mismatches:	6
Best Local Similarity:	13.6%	Indels:	0
Query Match:	3	Gaps:	0
DB:			

US-10-718-311-16 (1-170) x US-09-543-681A-3086 (1-195)

[illegible]

RESULT 15

```

US-09-818-780-75
; Sequence 75, Application US/09818780
; Patent No. 6677146
; GENERAL INFORMATION:
; APPLICANT: McHenry, Charles
; TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III HOLOENZYME
; FILE REFERENCE: 1794.0030004
; CURRENT APPLICATION NUMBER: US/09/818.780
; CURRENT FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/192,736
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 75
; LENGTH: 2127
; TYPE: DNA
; ORGANISM: Thermus thermophilus
US-09-818-780-75

```

Alignment Scores:			
Pred. NO.:	0.875	Length:	2127
Score:	50.00	Matches:	59
Percent Similarity:	32.1%	Conservative:	22
Best Local Similarity:	23.4%	Mismatches:	77
Query Match:	10.1%	Indels:	94
DB:	3	Gaps:	10

US-10-718-311-16 (1-170) x US-09-818-780-75 (1-2127)

Qy	9	ProAlaLeuThrGlnLeuArgAla-----LeuArgTyrCysLysGlu	22
Db	1027	CCCCGCTTGGCGAGCTTAGGGCCGGCTCCTCGGGTGGCCCTCACTTACTCTCCCTG	1086
Qy	23	IleProAlaLeuAsp-----ProGlnLeuLeuAspTrp	33
Db	1087	GTGCCCCGGTGGAGGTTTGGAGCGAAAGGCTTCGCCCTGCCCCGACCCCGC	1146
Qy	34	LeuLeuLeuGluAsp-----Ser	39

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2006, 03:03:17 ; Search time 1170.21 Seconds
(without alignments)
2677.596 Million cell updates/sec

Title: US-10-718-311-16
Perfect score: 887
Sequence: 1 MQVWMSHPALQRLALRYC.....RLSGKPLLLTELFLPASPLY 170

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-Q-/abs/ABSWEB_spool/US10718311/runat_26052006_164933_13735/app_query.fasta 1
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=rnpbm -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs05h
-USER=US10718311 @CGN_1_1 2326 @runat_26052006_164933_13735 -NCPU=6 -ICPU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA_Main:
1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq.*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq.*
15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq.*
16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	887	100.0	684	3	US-09-855-341-7 Sequence 7, Appli

2	887	100.0	684	3	US-09-896-866B-16	Sequence 16, Appl
3	887	100.0	684	7	US-10-359-369-41	Sequence 41, Appl
4	887	100.0	684	7	US-10-699-050-16	Sequence 16, Appl
5	887	100.0	684	8	US-10-718-311-7	Sequence 7, Appli
6	887	100.0	684	9	US-10-462-162-33	Sequence 33, Appl
7	862	97.2	3452	9	US-10-462-162-47	Sequence 47, Appl
8	854	96.3	495	3	US-09-855-341-3	Sequence 3, Appli
9	854	96.3	495	3	US-09-896-866B-12	Sequence 12, Appl
10	854	96.3	495	7	US-10-359-369-37	Sequence 37, Appl
11	854	96.3	495	8	US-10-699-050-12	Sequence 12, Appl
12	854	96.3	495	8	US-10-718-311-3	Sequence 3, Appli
13	854	96.3	498	9	US-10-462-162-29	Sequence 29, Appl
14	851	95.9	1971	9	US-10-462-162-54	Sequence 54, Appl
15	811	91.4	1207	10	US-10-450-763-26139	Sequence 26139, A
16	171.5	19.3	269233	8	US-10-672-787-41	Sequence 41, Appl
17	97.5	11.0	475	4	US-10-275-026A-113	Sequence 113, App
18	91	10.3	606	4	US-09-925-065A-789463	Sequence 789463,
19	91	10.3	606	5	US-09-925-065A-789463	Sequence 789463,
20	90	10.1	2127	9	US-10-746-167-75	Sequence 75, Appl
21	90	10.1	9133	10	US-10-915-740A-277	Sequence 277, App
22	90	10.1	2242716	10	US-10-915-740A-1068	Sequence 1068, Ap
23	88.5	10.0	2892	16	US-11-183-136-15	Sequence 15, Appl
24	88	9.9	42811	11	US-10-330-773-925	Sequence 925, App
25	87	9.8	59589	6	US-10-087-192-1315	Sequence 1315, Ap
26	86	9.7	522	10	US-10-467-657-2855	Sequence 2855, Ap
27	85.5	9.6	2907	7	US-10-108-260A-1067	Sequence 1067, Ap
28	85	9.6	888	10	US-10-617-320-1721	Sequence 1721, Ap
29	84.5	9.5	1780	10	US-10-487-078-92	Sequence 92, Appl
30	84.5	9.5	11331	10	US-10-461-862-101	Sequence 101, App
31	84	9.5	25001	7	US-10-017-161-2063	Sequence 2063, Ap
32	84	9.5	25001	7	US-10-292-798-1709	Sequence 1709, Ap
33	83.5	9.4	1514	8	US-10-424-599-27291	Sequence 27291, A
34	83.5	9.4	1565	10	US-10-750-185-33285	Sequence 33285, A
35	83.5	9.4	1365	10	US-10-750-623-33285	Sequence 33285, A
36	83.5	9.4	10710	8	US-10-672-396-4	Sequence 4, Appli
37	83.5	9.4	196200	15	US-11-121-086-10	Sequence 9, Appli
38	83.5	9.4	199321	15	US-11-121-086-10	Sequence 10, Appl
39	83	9.4	608	4	US-09-925-065A-772942	Sequence 772942,
40	83	9.4	608	5	US-09-925-065A-772942	Sequence 772942,
41	83	9.4	1140	3	US-09-928-175-22	Sequence 22, Appl
42	83	9.4	2314	3	US-09-928-175-19	Sequence 19, Appl
43	83	9.4	2337	8	US-10-437-963-39118	Sequence 39118, A
44	83	9.4	2539	6	US-10-229-735-4	Sequence 4, Appli
45	82.5	9.3	689	13	US-11-097-143-18275	Sequence 18275, A

ALIGNMENTS

RESULT 1
US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. US20020002715A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUD
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7

Alignment Scores:
Pred. No.: 9.84e-106 Length: 684

Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-855-341-7 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAACGCAACTCGTGCCTCGCTATTGT 231
Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTCGTGGAGATTCCATG 291
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACNAAACGTTTGAACACAGCGGAAACCGGTAGCGTGACGATGATCCGCGNAGGGTTT 351
Qy 61 ValGluGlnAnGluIleProGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGATGAATCCCGAAGAACTGCGCTGCTGCGCAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTTGTATGTCCGATGCTGAACCGTGGCTTGGCGGTCTGCTACCGTCTGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGCGCGAGCTGGCTTACAAAATTTGGGTAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCGACATTAACCGCGACTTTATTGAGATAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThr 160
Db 592 GCCGGGCTGTGGGGCGAGCTTCCGCTCGCATTAACCGGTAAACCGCTGTTCCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACGTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 2
US-09-896-866B-16
; Sequence 16, Application US/09896866B
; Patent No. US20020151002A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BCI034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein
US-09-896-866B-16

Alignment Scores: 9.84e-106 Length: 684
Pred. No.: 887.00 Matches: 170
Score: 887.00

Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-896-866B-16 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAACGCAACTCGTGCCTCGCTATTGT 231
Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTCGTGGAGATTCCATG 291
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACNAAACGTTTGAACACAGCGGAAACCGGTAGCGTGACGATGATCCGCGNAGGGTTT 351
Qy 61 ValGluGlnAnGluIleProGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGATGAATCCCGAAGAACTGCGCTGCTGCGCAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTTGTATGTCCGATGCTGAACCGTGGCTTGGCGGTCTGCTACCGTCTGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGCGCGAGCTGGCTTACAAAATTTGGGTAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCGACATTAACCGCGACTTTATTGAGATAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThr 160
Db 592 GCCGGGCTGTGGGGCGAGCTTCCGCTCGCATTAACCGGTAAACCGCTGTTCCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACGTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 3
US-10-359-369-41
; Sequence 41, Application US/10359369
; Publication No. US20030215927A1
; GENERAL INFORMATION:
; APPLICANT: E. I. duPont de Nemours and Company, Inc.
; APPLICANT: Viitanen, Paul
; APPLICANT: Meyer, Knut
; APPLICANT: Van Dyk, Drew
; TITLE OF INVENTION: UDP-Glucosyltransferases
; FILE REFERENCE: CL1821 US NA
; CURRENT APPLICATION NUMBER: US/10/359,369
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: 60/355,511
; PRIOR FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 41
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(684)
US-10-359-369-41

Alignment Scores: 9.84e-106 Length: 684
Pred. No.: 887.00 Matches: 170
Score: 887.00
Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 7 Gaps: 0
US-10-718-311-16 (1-170) x US-10-359-369-41 (1-684)
QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAAACCAACTGCGTGGCTGCGCTATTGT 231
QY 21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGTGCTGAGGATCCATG 291
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACATGATCCCGAAGGTTTT 351
QY 61 ValGluGlnAsnGluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCCGAGCAGAAATCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411
QY 81 LeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTGTATGTGCGCATGTGTAACCGTGGCTGTGCCGTCGTACCGTGT 471
QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CTTGTGTCAACGTTAAGCGGGCGGAGCTGGGTACAAAAATTTGGTTAAACGCGTTA 531
QY 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCATTAACCCGGGACTTTATTGAGATAGCCGCTGAT 591
QY 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCGGGCTGTGGGGCGAGCTTCCCGCTCGCATTAAGCGGTAAACCCGCTGTGTGCTAAC 651
QY 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GNACTGTTTTTACCGGCGTCAACGTTGTAC 681
RESULT 4
US-10-699-050-16
; Sequence 16, Application US/10699050
; Publication No. US20040142437A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/10/699,050
; CURRENT FILING DATE: 2003-10-30
; PRIOR APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein
US-10-699-050-16
Alignment Scores:
Pred. No.: 9,848-106 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Mismatches: 0

Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Mismatches: 0
DB: 8 Gaps: 0
US-10-718-311-16 (1-170) x US-10-699-050-16 (1-684)
QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAAACCAACTGCGTGGCTGCGCTATTGT 231
QY 21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGTGCTGAGGATCCATG 291
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACATGATCCCGAAGGTTTT 351
QY 61 ValGluGlnAsnGluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCCGAGCAGAAATCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411
QY 81 LeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTGTATGTGCGCATGTGTAACCGTGGCTGTGCCGTCGTACCGTGT 471
QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CTTGTGTCAACGTTAAGCGGGCGGAGCTGGGTACAAAAATTTGGGTAAACGCGTTA 531
QY 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCATTAACCCGGGACTTTATTGAGATAGCCGCTGAT 591
QY 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCGGGCTGTGGGGCGAGCTTCCCGCTCGCATTAAGCGGTAAACCCGCTGTGTGCTAAC 651
QY 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACTGTTTTTACCGGCGTCAACGTTGTAC 681
RESULT 5
US-10-718-311-7
; Sequence 7, Application US/10718311
; Publication No. US20040143867A1
; GENERAL INFORMATION:
; APPLICANT: Viitanen, Paul V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/10/718,311
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: US/09/855,341
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-10-718-311-7
Alignment Scores:
Pred. No.: 9,848-106 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Mismatches: 0

```

Query Match:      100.0%      Indels:      0
DB:               8          Gaps:        0

US-10-718-311-16 (1-170) x US-10-718-311-7 (1-684)

Qy      1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db      172 ATGAGGTGGGCATATGTACACCCCGCTTAACGCAACTGCGTGGCTCGCTATTGT 231
Qy      21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db      232 AAAGAGATCCCTGCGCTGGATCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATG 291
Qy      41 ThrLysArgPheGluGlnGlnGlySerThrValSerValThrMetIleArgGluGlyPhe 60
Db      292 ACAAAACGTTTTTGAACACGACGAGGAAAAACGGTAAGCGTGACGATGATCCCGAAGGGTTT 351
Qy      61 ValGluGlnAenGluIleProGluGluLeuProGluLeuProLysGluSerArgTyrTrp 80
Db      352 GTCGAGCAGAATGAATATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGG 411
Qy      81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db      412 TTACGTGAAATTTTGTATGTGCGCATGGTGAACCGTGGCTTGCCTGCTACCGTGGT 471
Qy      101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db      472 CCTGTGTCAACGTTAAGCGGGCGGAGCTGCCGCTTACAAAAATTTGGGTAAACCGCGTTA 531
Qy      121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db      532 GGACGCTATCTGTTACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGAT 591
Qy      141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db      592 GCCGGCTGTGGGGCGACGTTCCCGCTCGATTAAAGCGGTAACCGCTGTGCTAACCA 651
Qy      161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db      652 GAACTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 6
US-10-462-162-33
; Sequence 33, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-33

Alignment Scores:
Pred. No.:      9.84e-106      Length:      684
Score:          887.00      Matches:      170
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:      0
DB:             9          Gaps:        0

US-10-718-311-16 (1-170) x US-10-462-162-33 (1-684)

Qy      1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20

```

```

Db      172 ATGAGGTGGGCATATGTACACCCCGCTTAACGCAACTGCGTGGCTCGCTATTGT 231
Qy      21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db      232 AAAGAGATCCCTGCGCTGGATCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATG 291
Qy      41 ThrLysArgPheGluGlnGlnGlySerThrValSerValThrMetIleArgGluGlyPhe 60
Db      292 ACAAAACGTTTTTGAACACGACGAGGAAAAACGGTAAGCGTGACGATGATCCCGAAGGGTTT 351
Qy      61 ValGluGlnAenGluIleProGluGluLeuProGluLeuProLysGluSerArgTyrTrp 80
Db      352 GTCGAGCAGAATGAATATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGG 411
Qy      81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db      412 TTACGTGAAATTTTGTATGTGCGCATGGTGAACCGTGGCTTGCCTGCTACCGTGGT 471
Qy      101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db      472 CCTGTGTCAACGTTAAGCGGGCGGAGCTGCCGCTTACAAAAATTTGGGTAAACCGCGTTA 531
Qy      121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db      532 GGACGCTATCTGTTACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGAT 591
Qy      141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db      592 GCCGGCTGTGGGGCGACGTTCCCGCTCGATTAAAGCGGTAACCGCTGTGCTAACCA 651
Qy      161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db      652 GAACTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 7
US-10-462-162-47
; Sequence 47, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 47
; LENGTH: 3452
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of nucleic acid fragment encoding CPL (from
; OTHER INFORMATION: Escherichia coli), pRBA 1-hydroxylase (from Candida
; OTHER INFORMATION: parapsilosis), and UGT72B1 (from Arabidopsis thaliana).
US-10-462-162-47

Alignment Scores:
Pred. No.:      1.78e-101      Length:      3452
Score:          862.00      Matches:      166
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    97.2%      Indels:      0
DB:             9          Gaps:        0

US-10-718-311-16 (1-170) x US-10-462-162-47 (1-3452)

Qy      5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuPro 24
Db      1 CATATGTACACCCCGCGTTAAACGCAACTGCGTGGCTCGCTATTGTTAAAGAGATCCCT 60

```

Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	121	CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCCGAAGGGCTTTGTCGAGCAGAAATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACTCCCGCTGCTGCCGAAGAGTCTCGTTACTGTGTACGTGAAATTTTG	240
Qy	86	LeuCySAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCCGATGGTGAACCGTGGCTTCGCCGTCGTACCGTCTCTGTGTCAACGTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	301	AGCGGGCCGAGCTGGCGTTACAAAAATTGGGTAAAAACGCCGTTAGGACGCATATCTGTT	360
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGACATTAAACCCGGGACTTTATTGATAGTAGCCGGTATGCCGGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	421	CGACGTTCCCGCTCGGATTAAACGGGTAAAACCGCTGTTGCTAACAGACTGTTTTTACCG	480
Qy	166	AlaSerProLeuTyr	170
Db	481	CGGTACACCGTTGTAC	495

RESULT 9
 US-09-896-866B-12
 ; Sequence 12, Application US/09896866B
 ; Patent No. US20020151002A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Flint, Dennis
 ; APPLICANT: Meyer, Knut
 ; APPLICANT: Viitanen, Paul
 ; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate
 ; TITLE OF INVENTION: Benzoin Acid Glucosides
 ; FILE REFERENCE: BC1034 US NA
 ; CURRENT APPLICATION NUMBER: US/09/896,866B
 ; CURRENT FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: 60/216,615
 ; PRIOR FILING DATE: 2000-07-07
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 12
 ; LENGTH: 495
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 US-09-896-866B-12

Alignment Scores:			
Pred. No.:	1.35e-101	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatch:	0
Query Match:	96.3%	Indels:	0
DB:	3	Gaps:	0
US-10-718-311-16 (1-170) x US-09-896-866B-12 (1-495)			
Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTCACACCCCGGTTAAGCAACTGCTGCGCTATTGTTAAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuGlnLeuAspSerMetThrLysArgPheGlu	45
Db	61	CTGGATCCGCAACTGCCTCGACTGCTGTGCTGAGGATTCATGACAAAACGTTTTTGA	120
Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValcGlnAsnGlu	65
Db	121	CAGCAGGAAAAACGGTAAGCGTCACCATGATCCCGAAGCGTTGTTCGACGAGATGAA	180

```
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
Db 181 ATCCCGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGTGAACCGTGGCTTGCCTGCTACCGCTGTTCTGTGTCAACGTTA 300
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 301 AGCGGGCGAGCTGGCTTACAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGATGCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGAGGTTCCCGCCCTGCGATTAAAGCGGTAACCGCTGTGTCTTAACAGAACTGTTTACCG 480
Qy 166 AlaSerProLeuTyr 170
Db 481 CGGTACCGTTGTAC 495

RESULT 10
US-10-359-369-37
; Sequence 37, Application US/10359369
; Publication No. US20030215927A1
; GENERAL INFORMATION:
; APPLICANT: E. I. duPont de Nemours and Company, Inc.
; APPLICANT: Viitanen, Paul
; APPLICANT: Meyer, Knut
; APPLICANT: Van Dyk, Drew
; TITLE OF INVENTION: UDP-Glucosyltransferases
; FILE REFERENCE: CL1821 US NA
; CURRENT APPLICATION NUMBER: US/10/359,369
; CURRENT FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: 60/355,511
; PRIOR FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(495)
US-10-359-369-37
```

```
Alignment Scores:
Pred. No.: 1.35e-101 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 7 Gaps: 0
```

US-10-718-311-16 (1-170) x US-10-359-369-37 (1-495)

```
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
Db 1 ATGTACACCCCGCGTTAAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 60
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65
Db 121 CAGCAGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGGTTTGTGACAGCAATGA 180
```

```
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
Db 181 ATCCCGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGTGAACCGTGGCTTGCCTGCTACCGCTGTTCTGTGTCAACGTTA 300
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 301 AGCGGGCGAGCTGGCTTACAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGTGATGCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGAGGTTCCCGCCCTGCGATTAAAGCGGTAACCGCTGTGTCTTAACAGAACTGTTTACCG 480
Qy 166 AlaSerProLeuTyr 170
Db 481 CGGTACCGTTGTAC 495
```

RESULT 11

US-10-699-050-12

```
; Sequence 12, Application US/10699050
; Publication No. US20040142437A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/10/699,050
; CURRENT FILING DATE: 2003-10-30
; PRIOR APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-699-050-12
```

```
Alignment Scores:
Pred. No.: 1.35e-101 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 8 Gaps: 0
```

US-10-718-311-16 (1-170) x US-10-699-050-12 (1-495)

```
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
Db 1 ATGTACACCCCGCGTTAAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 60
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65
Db 121 CAGCAGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGGTTTGTGACAGCAATGA 180
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
```

Db	131	ATCCCCGAGAAGAACTGCCCGCTGCTGCCGAAGAAGTCTCGTTACTCGTGTTACGTGAAATTTTG	240
Qy	86	LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCCGATGGTGAACCGGTGGCTTGCCTGTCACCGTGGTCTGTGTCAACGTGA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnIlyValLeuGlyIlyeThrProLeuGlyArgTrIleuPhe	125
Db	301	AGCGSGCCGGAGCTGGCGTTACAAANNTTGGGTAAACCCGTTAGGACGCTATCTGTTC	360
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGACATTAAACCCGGACTTTATTGAGATAGGCGGTGATGCCGGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyIlyysProLeuLeuLeuThrCluLeuPheLeuPro	165
Db	421	CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTCCTAACGACTGTTTTTACC	480
Qy	166	AlaSerProLeuTyr	170
Db	481	GCCTCACCGTTGTAC	495

```

RESULT 12
US-10-718-311-3
; Sequence 3, Application US/10718311
; Publication No. US20040143867A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUIT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/10/718,311
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: US/09/855,341
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-10-718-311-3

```

Alignment Scores:		
Pred. NO.:	1.35e-101	Length: 495
Score:	854.00	Matches: 165
Percent Similarity:	100.0%	Conservative: 0
Best Local Similarity:	100.0%	Mismatches: 0
Query Match:	96.3%	Indels: 0
DB:	8	Gaps: 0

US-10-718-311-16 (1-170) x US-10-718-311-3 (1-495)

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTCACACCCCGCGTTAACGCAACTGCGTGGCTGCCTATTGTTAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu	45
Db	61	CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAAAGTTTGTAA	120
Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValIcLuGlnAsnGlu	65
Db	121	CAGCAGGGAAAAACGGTTAAGCGTGACGATATCCCGAAGGGTTTGTCCAGCAGAAATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACTGCCGCTGTGCGCAAGAAAGCTTCGTATTCTGTTACGTGAAATTTTG	240

Qy	86	LeuCysAlaAspGlyGluProTrpLeuAlaGlyAArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCGGATGTTGAACCGTGGCTTGC	300
Qy	106	SerGlyProGluLeuAlaLeuAlaLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	301	AGCGGGCCGGAGCTGGCGTTTCAAAANTTGGTAAACCCGTTAGAGCGCTATCTGTC	360
Qy	126	ThrSerSerThrLeuThrArgAspPheLeuGlyLeuGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGACATTAAACCGGACCTTATTAGATAGAGCCGCGTATCCCGGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	421	CGAGCTTCCCGCTCGGATTAAACGGTAAACCCGCTGTGCTAACAGACTGTTTTACCG	480
Qy	166	AlaSerProLeuTyr	170
Db	481	CGCTCACCCTTGATC	495

```

RESULT 13
US-10-462-162-29
; Sequence 29, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-29

```

Alignment Scores:		
Pred. No.:	1,36e-101	Length:
Score:	854.00	Matches:
Percent Similarity:	100.0%	Conservative:
Best Local Similarity:	100.0%	Mismatches:
Query Match:	96.3%	Indels:
DB:	9	Gaps:

US-10-718-311-16 (1-170) x US-10-462-162-29 (1-498)

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTCACACCCCGCGTTAAACCAACTGCTGCGCTGCCTATTGTTAAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu	45
Db	61	CTGATCCGCAACTGCTCGACTGCTGTTCGTGGAGGATTCATCACAAAACGTTTGGAA	120
Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	121	CAGCAGGGAAAAACGGTAAAGCGTGACGATGATCCCGCAAGGGTTTGTCCAGCAGCAATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACTCCCGCTCTGCCGAAGAGTCTCGTTACTGGTTACGTCAAAATTTTG	240
Qy	86	LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTCCCGATGGTGAACCGTGCTTCCCGGTCTGACCGTCGTTCTCTGTGTCAACGTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125

Db 301 AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 360
Qy 126 ThrSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGACGTTCCCGCCCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480
Qy 166 AlaSerProLeuTyr 170
Db 481 CCGTCACCGTTGTAC 495
RESULT 14
US-10-462-162-54
; Sequence 54, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 1971
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleic acid sequence of a nucleic acid fragment inserted into
; OTHER INFORMATION: expression vector pET29a encoding CPL (from Escherichia coli) and
; OTHER INFORMATION: pHEA 1-hydroxylase (from Cadida parapsilosis).
US-10-462-162-54
Alignment Scores:
Pred. No.: 2,27e-100 Length: 1971
Score: 851.00 Matches: 165
Percent Similarity: 99.4% Conservative: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 99.9% Indels: 0
DB: 9 Gaps: 0
US-10-718-311-16 (1-170) x US-10-462-162-54 (1-1971)
Qy 5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluilePro 24
Db 1 CATATGTACACCCCGCGTTAAACGCACTGCGTGCCTATTATTTAAAGAGATCCCT 60
Qy 25 AlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPhe 44
Db 61 CCCCTGGATCCGCAACTGCTGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTT 120
Qy 45 GluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsn 64
Db 121 GAACAGCAGGAGAAAACCGTAAAGCGTGACGATGATCCCGAAGGTTTGTCTGAGCAGAA 180
Qy 65 GluileProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluile 84
Db 181 GAAATCCCGAAGAACTCCGCTGCTGCCGGAAGAGTCTCGTTACTGTTACGTGAAAT 240
Qy 85 LeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThr 104
Db 241 TTGTTATGTCGATGGTGAACCGGTGGCTTGCCTGCTACCGTCTCTGTTACGTGAAAT 300
Qy 105 LeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeu 124
Db 301 TTAAGCGCGCGAGCTGGCGTTACAAAAATTGGGTAAACCGCGTTAGGACGCTATCTG 360

Qy 125 PheThrSerSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrp 144
Db 361 TTCATCATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGGCTGTGG 420
Qy 145 GlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeu 164
Db 421 GGGCGAGTTCGCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTA 480
Qy 165 ProAlaSerProLeuTyr 170
Db 481 CCGCGCTCACCGTTGTAC 498
RESULT 15
US-10-450-763-26139
; Sequence 26139, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 26139
; LENGTH: 1207
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (893)..(925)
; OTHER INFORMATION: 81% homologous to Escherichia coli 4-hydroxybenzoate
; OTHER INFORMATION: synthetase, accession number X57434, Smith-Waterman Score=49.
US-10-450-763-26139
Alignment Scores:
Pred. No.: 2,05e-95 Length: 1207
Score: 811.00 Matches: 163
Percent Similarity: 98.2% Conservative: 1
Best Local Similarity: 97.6% Mismatches: 1
Query Match: 91.4% Indels: 2
DB: 10 Gaps: 0
US-10-718-311-16 (1-170) x US-10-450-763-26139 (1-1207)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 25
Db 652 ATGTACACCCCGCGTTAAACGCACTGCGTGCCTATTGTTAAAGAGATCCCTGCC 711
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 712 CTGGATCCGCAACTGCTGCGTGGCTGCTGCTGAGGATTCATGACAAAACGTTTTGAA 771
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 772 CAGCAGGAGAAAACCGTAAAGCGTGACGATGATCCCGAAGGTTTGTCTGAGCAGAA 831
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 85
Db 832 ATCCCGAAGAACTGCCGCTGCTGCCGGAAGAGTCTCGTTACTGTTACGTGAAATTTTG 891
Qy 86 -LeuCysAlaAspGlyGluProTrpLeuAla-GlyArgThrValValProValSerThrL 105
Db 892 TGTATGTCGCGATGGCGAACCAGCGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
Qy 105 euserGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuP 125

```

Db      952 TAAGCGGCGCGAGCTGGCGTTACAAAAATTGGTAAAAACGCCGTTAGGACGCTATCTGT 1011
Qy      125 heThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 145
Db      1012 TCACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCGCGTGATGCCGGGCTGTGGG 1071
Qy      145 lyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuP 165
Db      1072 GCGGACGTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTAC 1131
Qy      165 roAlaSerProLeuTyr 170
Db      1132 CCGCGTCACCGTTGTAC 1148

```

Search completed: May 30, 2006, 03:41:56
 Job time : 1175.21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2006, 03:03:47 ; Search time 36.0299 Seconds
(without alignments)
798.935 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MOVHMSHPALTQLPALRYC.....RLSGKPLLLTELFLPASPLY 170

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 242596 seqs, 56442199 residues

Total number of hits satisfying chosen parameters: 485192

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/abs/ABSSWEB/spool/US10718311/runat_26052006_164936_13807/app_query.fasta_1
-DB=Published Applications NA New -GFMT=fastap -SUFFIX=rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-MAXLEN=2000000000 -HOST=abs03h
-USER=US10718311 @CGN 1.1.26 @runat_26052006_164936_13807 -NCPU=6 -ICPU=3
-NO MAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : Published Applications NA New.*

1: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US09_NEW_PUB.seq.*
2: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US08_NEW_PUB.seq.*
5: /EMC_Celerra_SIDS3/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
6: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US11_NEW_PUB.seq.*
8: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	9.6	2907	7	US-11-293-697-1067
2	80	9.0	3023	7	US-11-293-697-665
3	76.5	8.6	3550	7	US-11-293-697-509
4	76	8.6	1104	7	US-11-217-529-2348
5	75.5	8.5	601	6	US-10-488-619-1930
6	75.5	8.5	1791	7	US-11-217-529-1938
7	74.5	8.4	2384	7	US-11-293-697-1066
8	73	8.2	489	7	US-11-217-529-174332
9	73	8.2	1406	6	US-10-953-349-34874
Sequence 1629, Ap					
Sequence 1, Appli					
Sequence 200, App					
Sequence 1, Appli					
Sequence 336, App					
Sequence 48, Appl					
Sequence 4803, Ap					
Sequence 24, Appl					
Sequence 7735, A					
Sequence 2, Appli					
Sequence 1696, Ap					
Sequence 191165, A					
Sequence 76487, A					
GENERAL INFORMATI					
Sequence 1, Appli					
Sequence 173132, A					
Sequence 166866, A					
Sequence 33008, A					
Sequence 3713, Ap					
Sequence 76752, A					
Sequence 78541, A					
Sequence 21628, A					
Sequence 20563, A					
Sequence 596, App					
Sequence 40225, A					
Sequence 37006, A					
Sequence 22504, A					
Sequence 1, Appli					
Sequence 239, App					
Sequence 151, App					
Sequence 68, Appl					
Sequence 75461, A					
Sequence 51, Appl					
Sequence 29, Appl					
Sequence 76569, A					
Sequence 7, Appli					

ALIGNMENTS

RESULT 1

US-11-293-697-1067
; Sequence 1067, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1067
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1067

Alignment Scores:
Pred. No.: 1.37
Score: 85.50
Length: 2907
Matches: 50
Percent Similarity: 40.3%
Conservative: 21
Best Local Similarity: 28.4%
Mismatches: 67
Query Match: 9.6%
Indels: 38
Gaps: 10
DB:

US-10-718-311-16 (1-170) x US-11-293-697-1067 (1-2907)

QY 3 ValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGlu 22

DB 484 GTCTGGCCCTTACCCATCCCAAGCTG----- 510

```
QY 23 IleProAlaLeuAspProGlnLeuLeuAspTrp-----LeuLeuLeuGluAspSerMet 40
Db 511 ATCCCTACTAGGTCACAGAGCGTCTCCACCAGCAGCTCCAGAGGGATGATAAATTG 570
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArg-----Glu 58
Db 571 TCAAAAGCCACCAAGAGTCAGGACAGGAAGTCTAGTTCGCCAGCTGAGGGCTGTGCAG 630
QY 59 GlyPhe-----ValGluGlnAsnGlnIleProGluGluLeuProLeuLeuProLysGlu 76
Db 631 GGATTTCAGGTACCTGTGCGCAACCCAGCCCTCACTCAATTCAACTCAGTCAGAAACCA 690
QY 77 SerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpTrpLeuAlaGly 96
Db 691 CCGCTCAGTGGATTAGAAA-----TGTGACAGGGAT-----TTCAGAGGG 732
QY 97 ArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGly 116
Db 733 CAATGGGTCCAGCCC-----TCACACCAGCACCACCTCTGCAGGCATGGNAA 780
QY 117 LysThrProLeuGly-Arg-----TyrLeuPheThrSerSerThrLeuThrArgAspPh 134
Db 781 CGCAGGCCAGAGGACAGGGCTTCTCAATATGGGCACACAGCACCACCAAGACAGCTCCATA 840
QY 134 eilegluile-----GlyArgAspAlaGlyLeuTrp-----GlyAr 146
Db 841 AATAGAGTGAATATCCAAAGCTCCCACTCCAGGGAAGCTGACCACCTGGGTGGATGACTC 900
QY 146 GArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGlu 161
Db 901 CCCAGCGTCCAAACCCCTCTCTGGCAGTCCCTTGTGTGTCTCCAGGAG 946

RESULT 2
US-11-293-697-665
; Sequence 665, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NOVEL full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5456
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 665
; LENGTH: 3023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-665

Alignment Scores:
Pred. No.: 6.6 Length: 3023
Score: 80.00 Matches: 36
Percent Similarity: 39.1% Conservative: 16
Best Local Similarity: 27.1% Mismatches: 53
Query Match: 9.0% Indels: 28
DB: 7 Gaps: 5

US-10-718-311-16 (1-170) x US-11-293-697-665 (1-3023)
QY 5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIlePro 24
Db 445 CACTTAGAAGACCACATTATAGTGAGCTGACCAACCTGAAGGTGTGCATTGAATTAACA 504
QY 25 AlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPhe 44
Db 505 GGGCTCCATCTTAATAAACCAACGCCACTTGCTG-----CACCTTAGAAGACGATGG 555
QY 45 GluGlnGln-----GlyLysThrValSer 52
|||||
```

```
Db 556 GAGCAGCAGGTGTCCGCAGCAGATGGCAAAACCTGCGCGGCAAGCAGGAGGAAGTGACC 615
QY 53 ValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeuPro--- 71
Db 616 CAGGCCACTCAGCCTGAGGCCATTCTCCAGGGGACTAACATCACTGAAGAGAAACCTGGC 675
QY 72 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAla 88
Db 676 AGGAAAAGGCGAGAGGCCAAGGCCAAGCAAGTAGCTGGTCGGAAGAGTCTCTTAAACCCAGT 735
QY 89 AspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeuSerGlyPro 108
Db 736 GACAATGACAAGGCTTG-----CCTGTGTCTCCGGCTCTCCGCC 777
QY 109 GluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 121
Db 778 -----ATGAAGAGTCTTTCATCCACCAGTCAGGC 807

RESULT 3
US-11-293-697-509/c
; Sequence 509, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NOVEL full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 509
; LENGTH: 3550
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-509

Alignment Scores:
Pred. No.: 21.9 Length: 3550
Score: 76.50 Matches: 43
Percent Similarity: 41.1% Conservative: 15
Best Local Similarity: 30.5% Mismatches: 35
Query Match: 8.6% Indels: 49
DB: 7 Gaps: 8

US-10-718-311-16 (1-170) x US-11-293-697-509 (1-3550)
QY 32 AspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 51
Db 397 GACACACCTTAAGTGAAGAGTGCCTCACAATACAGAGTCCACAGAGGGGCAAT--GCG 340
QY 52 SerValThrMetIleArgGluGlyPheValGluGln-----Asn 64
Db 339 AGTGTG-----GAGGGAAGCATCTCCAGGATTCAGGGCGCCCGCTGACGAC 292
QY 65 GluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIle 84
Db 291 TCCTGCCGCGCAAAAGGTCCCGTGTCTCCACGAGAATGAGAGCATGGCCCTGACCCATC 232
QY 85 -----LeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValPro--- 101
Db 231 GTCAGGTTGACCTGC-----TGG-----CCAACCTCAGTTTCTCCAAGG 193
QY 102 ValSerThrLeuSerGlyProGluLeuAla----- 111
Db 192 GTCACATGCTCTTTGGTCTCTGAGTTACCCCTCATCGGAGACGAGAGGCATGGGAGGAAA 133
QY 112 -----LeuGlnLysLeuGlyLysThrProLeu 120
Db 132 AGGGGGTACACAGCCGCGAAGGAGTTGGCTTCAGACAGAAAAGACAAAGAACCCGGGTG 73
QY 121 GlyArgTyrLeuPhe-----ThrSerThrLeuThrArgAspPhe 134
```

```
Db 72 GGAAGATGGCTTCTGCGGGTAGGAATCACACACTTCGTCAACTGTGCAAAAGACAGA 13
QY 135 ile 135
Db 12 ATT 10
RESULT 4
US-11-217-529-2348
; Sequence 2348, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2348
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-2348
Alignment Scores:
Pred. No.: 4.62 Length: 1104
Score: 76.00 Matches: 42
Percent Similarity: 38.3% Conservative: 28
Best Local Similarity: 23.0% Mismatches: 63
Query Match: 8.6% Indels: 50
DB: 7 Gaps: 10
US-10-718-311-16 (1-170) x US-11-217-529-2348 (1-1104)
QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIle---Pro 24
Db 547 GTAGATAGACCCATAATTTCTTCAATCACCGGAGTAGCGTACTCTACAGAGTAAGTCG 606
QY 25 AlaLeuAppProGlnLeuLeuAspTrp-----LeuLeuLeuGlu 37
Db 607 TCACCTCGTAAATAGAATGATGACGCGTGCCAAAAAAGTATTGAAGGACCCTCCAATGTGAG 666
QY 38 AspSerMetThrLys-----ArgPheGluGlnGlnGlyValThrVal-----Ser 52
Db 667 GTTACATACACAGCAGATGTTGGAGAGCGGAAATTCAGGTAAAGAGTCCAGCGTGGGGT 726
QY 53 ValThrMetIleArgGluGly-----PheValGluGln-----63
Db 727 ATAACTTTGGTGGCTGAATCCCAAGCAGAAAAGTTGGAGCTATTTCGCCGAAGATATTGT 786
QY 64 -----AsnGluIleProGluGlnLeuProLeuProLysGluSerArgTyrTrpLeu 81
Db 787 GATGAGGCTCTATACCCGAGAACTTGGT-----GAAAAAGTTGCTTACCCTTTATTA 840
QY 82 ArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgValValPro 101
Db 841 GAAGAAATATCGAAAAGCGCT-----GCTGTCGGTAGAAGTCAACTTCCA 885
QY 102 ValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 121
Db 886 CTAGCA-----ATCGTATACATGGTCATCGGAAAGAGATATCGG 927
QY 122 ArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAla 141
Db 928 AGATTAAAGAAATTAATAAGGACCAGATAGATGAAGGTTTATAACCCCTCTTGAGAGATATA 987
```

```
QY 142 GlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGlu 161
Db 988 -----AAGAAGATCTTCTGCACAGAA 1008
QY 162 LeuPheLeu 164
Db 1009 GTCTTTTG 1017
RESULT 5
US-10-488-619-1930
; Sequence 1930, Application US/10488619
; Publication No. US20060099578A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P.C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; FILE REFERENCE: 98-01 WO
; CURRENT APPLICATION NUMBER: US/10/488,619
; CURRENT FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 3040
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1930
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-488-619-1930
Alignment Scores:
Pred. No.: 2.2 Length: 601
Score: 75.50 Matches: 32
Percent Similarity: 47.2% Conservative: 28
Best Local Similarity: 25.2% Mismatches: 38
Query Match: 8.5% Indels: 29
DB: 6 Gaps: 9
US-10-718-311-16 (1-170) x US-10-488-619-1930 (1-601)
QY 15 ArgAlaLeuArg-----TyrCysLysGluIleProAlaLeuAsp 27
Db 151 CGTTCCTTAAGGAAAGCAAGCCATCACAGAAGTGTATTGCCCACCATCAAGGTCCACACT 210
QY 28 ProGlnLeuLeuAsp-----TrpLeuLeuLeuGluAspSerMetThrLys 42
Db 211 GCTGAGCTTCTACCATACAGAGCCATGGCTTGGGCTCCAGCTTGTGATCATCTTTAAG 270
QY 43 ArgPheGluGlnGlnGlyLysThrValSer-----ValThrMetIleArgGluGly 59
Db 271 AATTTTAAAGGAAAGCAAAATCATCTCTGAATATGACATCACCTTGATTATGACT--- 327
QY 60 PheValGluGlnAsnGluIleProGluGluLeuProLeuProLysGluSerArgTyr 79
Db 328 TACATAGAGGAAATTAAGCTACAGAAAGCTGTTTCTGTAATTTGAAAGGTA----- 378
QY 80 TrpLeuArgGluIle-----LeuLeuCysAlaAspGlyGluProTrpLeu 94
Db 379 ---CTGAGAGACATCGAGAGTCTCTCTGACATAGCTGTGCACAGGGGAAACAGGC--- 432
QY 95 AlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLys 114
Db 433 GCAGGGAAGTCCACTTTC---ATCAATACCTCTGAGGGG-----GTGGGGCATGAAGAA 483
QY 115 LeuGlyLysThrProLeuGly 121
Db 484 AAAGGTGACGCCCCCACTGGC 504
RESULT 6
US-11-217-529-1938
; Sequence 1938, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
```

```

; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1938
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1938

Alignment Scores:
Pred. No.:      10.7      Length:      1791
Score:          75.50     Matches:      36
Percent Similarity: 34.3%  Conservative: 10
Best Local Similarity: 26.9%  Mismatches:  41
Query Match:    8.5%      Indels:      47
DB:             7        Gaps:         8

US-10-718-311-16 (1-170) x US-11-217-529-1938 (1-1791)
Qy 67 ProGluGluLeuPro-----LeuLeuPro 74
Db |||:|||||
94 CCAGTAGCTTCCGCGAGCTGTGAAGCTGTTAAGAACACCGGTTGGCAGCTGGAGCCC 153
Qy 75 LysGluSerArgTyrTrpLeuArgGluLeuLeuCysAlaAspGlyGluProTrpLeu 94
Db |||:|||||
154 GCATTGACGAGTATTTCT-----GACGGGAA-----TGGAGG 186
Qy 95 AlaGly-----ArgThrValProValSerThrLeu 105
Db |||:|||||
187 GATGCTCCGAGGACGGATGGAGAGCCACTGCCACAGCCAGCCCTTCCACAGCCC 246
Qy 106 SerGlyProGluLeuAlaLeuGlnLys-----LeuGlyLysThrProLeuGlyArgTyr 123
Db |||:|||||
247 ATGGCAGAGGAGTTGGCTGGACTCCGCCCGTGTGGTCCAGGCCCTCTATCA----- 300
Qy 124 LeuPheThrSer-----SerThrLeuThrArgAspPhe 134
Db |||:|||||
301 ---TTTACAGCTGCATTCCCGTAGTCAGACCATTTGCCAGCGAGCTCCGTCATGATTTT 357
Qy 135 IleGluIle-----GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArg 151
Db |||:|||||
358 AGAACCATTTGGCTGAACGGAGCTCCCAATACCGTATGCTCATGTTCCAGTCTTCAGT 417
Qy 152 LeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db |||:|||||
418 TACAACGGTAACCCCTTTCTTTTGTGTTGCTGCTGATCCCG 459

RESULT 7
US-11-293-697-1066
; Sequence 1066, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1066
; LENGTH: 2384
; TYPE: DNA

; ORGANISM: Homo sapiens
US-11-293-697-1066

Alignment Scores:
Pred. No.:      21.3      Length:      2384
Score:          74.50     Matches:      25
Percent Similarity: 48.3%  Conservative:  4
Best Local Similarity: 41.7%  Mismatches:  21
Query Match:    8.4%      Indels:      10
DB:             2        Gaps:         2

US-10-718-311-16 (1-170) x US-11-293-697-1066 (1-2384)
Qy 70 LeuProLeu-LeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAs 89
Db |||:|||||
11 TTGCCACTCCCTTGGCAAGGAGAGCCGAGACCTCAGTTCCCGGGGCTCTTTGCCGGGGCAC 70
Qy 89 pGlyGluProTrpLeuAlaGlyArgThrValValPro---ValSerThrLeuSerGlyPr 108
Db |||:|||||
71 AGTGAGCCCTGGCTGCGCGCGGCCCTCTCCCGGGCCTCCACAGATGGGGGCTCC 130
Qy 108 oGluLeuAla-----LeuGlnLysLeuGlyLysThrPro 119
Db |||:|||||
131 GGAGGTGGCGCCAGGCTCTGAGCTACCTAGGTCTGCAGACTAGCGGGCATTTGGCCA 188

RESULT 8
US-11-217-529-174332
; Sequence 174332, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 174332
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-174332

Alignment Scores:
Pred. No.:      3.25      Length:      489
Score:          73.00     Matches:      39
Percent Similarity: 34.3%  Conservative:  7
Best Local Similarity: 29.1%  Mismatches:  56
Query Match:    8.2%      Indels:      32
DB:             7        Gaps:         6

US-10-718-311-16 (1-170) x US-11-217-529-174332 (1-489)
Qy 56 IleArgGluGlyPheValGluGlnAsnGluLeuProGluGluLeuPro----- 71
Db |||:|||||
31 TTAATGAGTTCAGGTCATCAGCAATTTCCCCCAAGACAGATTGCCAGATGGTGAGA 90
Qy 72 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGlu 83
Db |||:|||||
91 CTGTTGAGGAACCATGTTGGCAGTTAGAACACGACATTGAGCCGATATTTCCATGGAGAG 150
Qy 84 IleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 103
Db |||:|||||
151 -----TGGAAAGCGCAACACGACAGATGGGAGAGTCC---ACTCAACAGTCT 195
Qy 104 ThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyr 123

```

```
Db 196 ACACCCATGGCTGAGACCTTGGTTCTCCCGCTTGGGACCTAGACCGTTG----- 246
QY 124 LeuPheThrSerSer-----ThrLeuThrArgAspPhe 134
Db 247 TTTATTTACAGCTCTCACTGCCGCTAGTCAGACCATTCGCCAGCGAAGTTCGCAATGATTTC 306
QY 135 IleGluIle-----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArg 151
Db 307 AGAACAACTGGTTTAAATGAGCGCTCTAACACCTGATGTTCCATGTTTGAATCTTTTACT 366
QY 152 LeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 367 TATGATGGTAACCCCTTCTTTTATTATTACTGTTGATCCCT 408

RESULT 9
US-10-953-349-34874/c
; Sequence 34874, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERY
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 34874
; LENGTH: 1406
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-34874
```

```
Alignment Scores:
Pred. No.: 15 Length: 1406
Score: 73.00 Matches: 29
Percent Similarity: 42.0% Conservative: 21
Best Local Similarity: 24.4% Mismatches: 35
Query Match: 8.2% Indels: 34
DB: 4 Gaps: 4

US-10-718-311-16 (1-170) x US-10-953-349-34874 (1-1406)
QY 56 IleArgGluGlyPheValGluGlnAsnGluIlePro-----GluGlu 69
Db 906 CTGAGTCGACGACCGTAGCTTGTCCAGCTTGCCGATCTCCGGCGGGATGCGAGGAGG 847
QY 70 LeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAsp 89
Db 846 ATGCCGCTGTTGAGGAGG----- 829
QY 90 GlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGlu 109
Db 828 -----ATGAGCTCGTTGAGCTGCGCGGCC 805
QY 110 LeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerThr 129
Db 804 ATGTGCCCATGCTGCTCGGAGCGCGCTCACCCGGAAGTTTC---GCCAGCAGC 748
QY 130 LeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyArgSerArg 149
Db 747 ATCACC---GACGCCCTCAGTTGGTAGCTGGCGGCGAGCGTGAACCTCGAAGTGGTTG 691
QY 150 LeuArgLeuSerGlyLysProLeuLeuLeuLeuThrGluLeuPheLeuProAlaSerPro 168
Db 690 TTGTTGATGAAAGCGCGTCGATCTTCTTGTCAAGATGCGCGCGCCGACCTCGCG 634

RESULT 10
US-11-293-697-1629/c
; Sequence 1629, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
```

```
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1629
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1629

Alignment Scores:
Pred. No.: 25.4 Length: 2019
Score: 73.00 Matches: 51
Percent Similarity: 33.2% Conservative: 20
Best Local Similarity: 23.8% Mismatches: 49
Query Match: 8.2% Indels: 94
DB: 9 Gaps: 9

US-10-718-311-16 (1-170) x US-11-293-697-1629 (1-2019)
QY 1 MetGlnValTrpHisMetSerHisPro-----AlaLeuThrGlnLeu 14
Db 1968 TTAAGATGTGGAAGATGGGTACAAATTAACCATGAGAGTTGTGCAGGGAACACCCGT 1909
QY 15 ArgAlaLeuArgTyrCysLysGluIleProAla-----LeuAspProGlnLeuLeu 31
Db 1908 AGGGCTGTTTGCACCTTCAGATATTCCTGCTCCCAAAATTCAGACCCCCAGATG--- 1852
QY 32 AspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 51
Db 1851 -----CAGGGCAAGACA--- 1840
QY 52 SerValThrMetIleArgGluGlyPheValGluGlnAsnGluIlePro----- 67
Db 1839 -----ATAAGAAGGGTGAGTCAAGCAAGGAGAGACCTCTCTGTAAGAGGCTG 1792
QY 67 ----- 67
Db 1791 AGGTCCCTCTGTTTCAAGGATGGATGTCAAGCTTGACCTTCCGGGGTCTGCAGTGGC 1732
QY 68 GluGluLeuProLeuLeuProLysGluSerArgTyrTrp---LeuArgGluIleLeuLeu 86
Db 1731 CAGAGGCTGCCGTGCGCCCTTCTCTCCCTTCTTGGGCACCTGTGGGAGCTTCTGGGT 1672
QY 87 CysAlaAspGly-GluProTrpLeuAlaGlyArgThrValProValSerThrLeuSe 106
Db 1671 CCTGCTGGNGGTGGTCCC----- 1653
QY 106 rGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro-LeuGlyArgTyrLeuPhe 126
Db 1652 -----CTCAGGCCGTGGGTTCAGTCTCTCTAGGAAGTCTCTCTTTG 1609
QY 126 hrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyA 146
Db 1608 CC-----CTTCGTGTCCTGGAAGGGGCTTGTCTGGAGGCA 1573
QY 146 rArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeu---- 164
Db 1572 AAGCGTCTCCACTCTGTCTCCTCAGACTCAGCTGTGTGGCCTTGATTTCTTTTGGCGG 1513
QY 165 -----ProAlaSerPro 168
Db 1512 ACTTGGCCCTTGGGTGCCAACGGTCCAGGATCCCCC 1475

RESULT 11
US-10-524-433-1
; Sequence 1, Application US/10524433
; Publication No. US20060099676A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Functional Genetics, Inc.
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN RAPAMYCIN
; FILE REFERENCE: 10784-023-228
; CURRENT APPLICATION NUMBER: US/10/524,433
; CURRENT FILING DATE: 2005-02-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 966
; TYPE: DNA
; ORGANISM: murine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 548, 564, 567, 583, 603, 610, 636, 647, 648, 651, 654, 656,
; LOCATION: 661, 667, 669, 675, 691, 710, 713, 719, 732, 746, 749, 754,
; LOCATION: 761, 771, 777, 781, 792, 795, 802, 808, 812, 813, 824, 833,
; LOCATION: 841, 842, 856, 861, 866, 871, 873, 874, 875, 882, 887
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 897, 902, 905, 910, 915, 923, 927, 928, 935, 938, 948, 949,
; LOCATION: 965, 966
; OTHER INFORMATION: n = A,T,C or G
US-10-524-433-1

Alignment Scores:
Pred. No.: 13.2      Length: 966
Score: 71.50        Matches: 33
Percent Similarity: 40.8%      Conservative: 18
Best Local Similarity: 26.4%    Mismatches: 56
Query Match: 8.1%             Indels: 19
DB: 6                   Gaps: 4

US-10-718-311-16 (1-170) x US-10-524-433-1 (1-966)
Qy 44 PheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGln 63
Db 385 TTTAAACACCAAGGAAAG---TACACAGCCTCCCGCCTCAGCGGAGCCCAAGT 441
Qy 64 AsnGluIleProGluLeuProLeuGluProLysGluSerArgTyrTrpLeuArgGlu 83
Db 442 AGCCACCTTCTTGCATCTGCCACACCTGCCACAGGAGACCTGCACCTGCCCGTGG 501
Qy 84 IleLeuLeuCyAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 103
Db 502 GCTTTGG-TGC-----TGGCTCTGTACTCCACTTGTCTGCTGTGGCC 544
Qy 104 ThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyr 123
Db 545 TCANGGAAAGT---CATCTCTNAANGAAACAAAGCCACCTNCTCCCTCCCTGGCATG 601
Qy 124 LeuPheThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeu 143
Db 602 CNTTTTCNGCCACAGCCCGTGTGTCCACANAAATTTGGCNCANGANGNAGGN 661
Qy 144 TrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPhe 163
Db 662 TGGGANAAACCCNAAACTA-----TTGTGG 688
Qy 164 LeuProAlaSerPro 168
Db 689 CTNCCCTGTGCCCC 703

RESULT 12
US-11-293-697-200/c
; Sequence 200, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106

```

```

; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 200
; LENGTH: 2490
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-200

Alignment Scores:
Pred. No.: 52      Length: 2490
Score: 71.50      Matches: 34
Percent Similarity: 45.4%      Conservative: 15
Best Local Similarity: 31.5%    Mismatches: 30
Query Match: 8.1%             Indels: 29
DB: 7                   Gaps: 9

US-10-718-311-16 (1-170) x US-11-293-697-200 (1-2490)
Qy 27 AspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGluGln 46
Db 755 GACCTGGATAGGTAATTGGCTCTCTCTGCC-----TCAGAATCCCTTCAGCAG 705
Qy 47 GlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIle 66
Db 704 CAGGACCTCACATG-----GGCTTTGTG-----GTC 678
Qy 67 ProGluGlu-----LeuProLeu-----LeuProLysGluSerArgTyrTrpLeuArgGlu 83
Db 677 CCTGAGGAACCTCCGTTCTGTGGGTTCGACATCCAGATGGCTGGTGGCGGTGGAG 618
Qy 84 IleLeuLeuCyAlaAspGlyGluPro-----TrpLeuAlaGlyArgThrVal 99
Db 617 CTGGGGATCCACACATCTTCTCAGCCACCTGGCAATTAATGGAGGCTGAGAAGTCCCTG 558
Qy 100 -----ValProValSerThrLeuSerGlyProGluLeuAlaLeuGln 113
Db 557 AGAGACTATTTCGAAGAGGTGCCAGTAGCGTGGCTCAGT---CCAAGTCTCGAAGCCTCA 501
Qy 114 LysLeuGlyLysThrProLeuGly 121
Db 500 GAACCAAGG---ACGCCACTGGGA 480

RESULT 13
US-10-473-691B-1/c
; Sequence 1, Application US/10473691B
; Publication No. US20060099202A1
; GENERAL INFORMATION:
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Soltau, Daniel A.
; TITLE OF INVENTION: IMMUNOGLOBULIN CONSTRUCT CONTAINING TUMOR SPECIFIC p53BP2 SEQUEN
; TITLE OF INVENTION: FOR ELICITING AN ANTI-TUMOR RESPONSE
; FILE REFERENCE: 02755/100K313-US1
; CURRENT APPLICATION NUMBER: US/10/473,691B
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US02/10224
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: US 60/280,733
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4534
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (757) ..(3774)
; OTHER INFORMATION:
; PUBLICATION INFORMATION:

```

; AUTHORS: Naumovski, L. and Cleary, M.L.
 ; TITLE: The P53-binding protein 53BP2 also interacts with Bcl2 and impedes
 ; TITLE: cell cycle progression at G2/M
 ; JOURNAL: Mol. Cel. Biol.
 ; VOLUME: 16
 ; ISSUE: 7
 ; PAGES: 3884-3892
 ; DATE: 1996-07
 ; DATABASE ACCESSION NUMBER: Genbank / U58334
 ; DATABASE ENTRY DATE: 1996-07-02
 US-10-473-691B-1

Alignment Scores:
 Pred. No.: 124 Length: 4534
 Score: 71.50 Matches: 28
 Percent Similarity: 39.2% Conservative: 12
 Best Local Similarity: 27.5% Mismatches: 33
 Query Match: 8.1% Indels: 29
 DB: 6 Gaps: 5

US-10-718-311-16 (1-170) x US-10-473-691B-1 (1-4534)

QY 68 GluGluLeuProLeuLeuProLysGluSerArgTyr----- 79
 Db 1670 GAAGCAGAGCCCTGGCTGGGAAAGATCTGCATTGGAAGGACTCCAATCAGGGCCAACT 1611
 QY 80 ---TrpLeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThr 98
 Db 1610 GGAATGATTTAGACCTTAGTTTGTGTGAGCAGCCCGCCAGATCTCATGTGGGCAGTGTCT 1551
 QY 99 ValValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThr 118
 Db 1550 TGTATTTTCATC-----GGCCCTCTGAAGCC-----TGAATGACCAAGGAA 1509
 QY 119 ProLeuGlyArgTyrLeuPheThrSerThr-----LeuThr 131
 Db 1508 CCATCCGCGAGGCTGGCTTACCAGCAATTCAGGCCTTGAGGCGATCCGAGGCATAGTA 1449
 QY 132 ArgAspPheIle-----GluIleGlyArgAspAlaGlyLeuTrp 144
 Db 1448 GATGACTGGATATAGGACCTACTGCAGCCACACGCTTGGGGCTGACCGCGCTTGCTGG 1389
 QY 145 GlyArg 146
 Db 1388 GGAAGA 1383

RESULT 14

; Sequence 336, Application US/11293697
 ; Publication No. US20060105376A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: Novel full length cDNA
 ; FILE REFERENCE: HI-A0106
 ; CURRENT APPLICATION NUMBER: US/11/293,697
 ; CURRENT FILING DATE: 2005-12-05
 ; PRIOR APPLICATION NUMBER: US/10/108,260
 ; PRIOR FILING DATE: 2002-03-28
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 336
 ; LENGTH: 3105
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-11-293-697-336

Alignment Scores:
 Pred. No.: 82.2 Length: 3105
 Score: 71.00 Matches: 32
 Percent Similarity: 43.6% Conservative: 9
 Best Local Similarity: 34.0% Mismatches: 35
 Query Match: 8.0% Indels: 18
 DB: 7 Gaps: 4

US-10-718-311-16 (1-170) x US-11-293-697-336 (1-3105)

QY 43 ArgPheGluGln-----GlnGlyLysThrValSerValThrMetIleArgGlu 58
 Db 1933 CGTTTTGACCAGATCAATGCTGTGGATGAACATGTAATGTCAGAAAGGAAAGAG 1992
 QY 59 GlyPheValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArg 78
 Db 1993 AATTTTGGCCATCAAAATCCAGCTCAAGTGAACCT-----AAAGAGCTTATA 2040
 QY 79 TyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArg--- 97
 Db 2041 TATTTTCATGAAGCTGATCCTTTTGTGTGTGTCTCC-----TGGTGTAGGAGAGAAAA 2094
 QY 98 -----ThrValValProValSerThrLeuSerGlyProGluLeu 110
 Db 2095 AAGCTCTATGAAAGAAATATAGGAAGTTCTCTCTTTTCACACCTTATTTCAATTGACTGCTG 2154
 QY 111 AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeu 124
 Db 2155 GCTGCTTTAAAAAAGGATCCATTGTACCGTTGCTTT 2196

RESULT 15

; Sequence 48, Application US/10488619
 ; Publication No. US20060099578A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Greenlee, Wanner and Sullivan, P.C.
 ; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
 ; FILE REFERENCE: Physiological Conditions, And Genotyping Arrays
 ; CURRENT APPLICATION NUMBER: US/10/488,619
 ; CURRENT FILING DATE: 2004-03-01
 ; NUMBER OF SEQ ID NOS: 3040
 ; SOFTWARE: Patent In version 3.1
 ; SEQ ID NO 48
 ; LENGTH: 674
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(674)
 ; OTHER INFORMATION: n is g, c, a or t
 US-10-488-619-48

Alignment Scores:

Pred. No.: 10.3 Length: 674
 Score: 70.50 Matches: 53
 Percent Similarity: 37.4% Conservative: 14
 Best Local Similarity: 29.6% Mismatches: 59
 Query Match: 7.9% Indels: 53
 DB: 6 Gaps: 10

US-10-718-311-16 (1-170) x US-10-488-619-48 (1-674)

QY 2 GlnValTrpHisMetSerHisPro-AlaLeuThrGlnLeuArgAlaLeuArgTyrCysIly 21
 Db 498 CAGCGCTGGCTACCGCAGCCATCCAGCCCTGCTCAGCAGCCGAGGGGTCGG----- 447
 QY 21 sGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuLeuLeuAspSerMetTh 41
 Db 446 -----CCAGTCTCTTTCTCAGACTGGGAGAGCTGGATGCCGAGGAGGT 403
 QY 41 rLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVa 61
 Db 402 GCGCCGGGGCCAGGCGCACGGGAGGCC-----AGGAGAGAGCTGTGT 361
 QY 61 lGluGlnAsnGluIle-----ProGluGluLeuProle 72
 Db 360 GGATCTCTCAGGAGATGCTGCGCTCTCTGGGCCACTGAGCCAGCCCGCCGCCCC 301
 QY 72 uLeuProLys-----GluSerArgTyrTrpLeuArgGlu-IleLeuLeuCys 87

```

Db      300 AGCAGGGAAGGATGAGTGTGGGAGGGNAGGGCTGGTCCGTCTGATGGGACTTTGC 241
Qy      88 -----AlaAsp---GlyGluProTyrLeuAlaGlyArgThrValValProValSerThr 104
Db      240 ACCTCTGCTGATCCCGCGCGCCCTGGCTTGGAGGCTTGGCTGTCTCTCCAGCGTCT--- 184
Qy      105 LeuSerGlyProGluLeuAlaLeu-GlnLysLeuGlyLysThrProLeuGlyArgTyrLe 124
Db      183 -----CTCCTCCCTCCTGGGGAAGGTGCGCCCTTGGCGCGCAAGGT 145
Qy      124 u-----PheThrSerSerThrLeuThrArgAspPheIleGluIleGly----- 138
Db      144 TTTAGCTTTCAGCAACTGAGTAACCTTAGGGACAGGTGGAGGTGTGGGCCCGATCTAAC 85
Qy      139 -----ArgAspAlaGlyLeuTyrGlyArgArgSerArgLeu 150
Db      84 CCTTACCCATCTCTACTGTGACTGTGG---AGGGTCACCAAGTTG 39
```

Search completed: May 30, 2006, 05:54:13
Job time : 42.0299 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2006, 01:43:54 ; Search time 4389.04 Seconds
(without alignments)
3248.869 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVWMSPALQTLRALRYC.....RLSGKPLLTLELFPASPLY 170

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh

-O=/abs/ABSSWEB spool/US10718311/runat_26052006_164919_13483/app_query.fasta_1
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02h
-USER=US10718311@CGN 1.1.7986 @runat_26052006_164919_13483 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: gb_est1:*
2: gb_est3:*
3: gb_est4:*
4: gb_est5:*
5: gb_est6:*
6: gb_hic:*
7: gb_est2:*
8: gb_est7:*
9: gb_est8:*
10: gb_est9:*
11: gb_gss1:*
12: gb_gss2:*
13: gb_gss3:*
14: gb_gss4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	618	69.7	832	13	CL666350 PRI0152b_
2	508	57.3	828	13	CL687382 PRI0146c_
3	262	29.5	536	14	AG192190 Pan trogl
4	236	26.6	533	14	AG192116 Pan trogl

5	209	23.6	515	12	CC048367
6	183	20.6	563	13	CL650764
7	171	19.3	838	13	CL688660
8	160	18.0	440	14	AG192305
9	150	16.9	1110	11	BZ551713
C 10	138	15.6	862	11	BZ564576
C 11	118	13.3	739	14	AJ863670
C 12	105	11.8	1141	12	BZ578851
C 13	95	10.7	753	13	CL696699
C 14	93	10.5	581	2	BG799182
15	92.5	10.4	4527	6	CR861280
16	92	10.4	1665	13	CW917940
17	91	10.3	666	2	BJ932163
C 18	91	10.3	1033	14	DU782699
C 19	90.5	10.2	668	10	DY396666
20	90	10.1	826	8	CO384679
C 21	90	10.1	1888	6	AY850332
C 22	89	10.0	451	3	BM870249
C 23	89	10.0	528	12	CE278012
24	89	10.0	770	7	BE618275
C 25	88.5	10.0	674	8	CX288046
C 26	88.5	10.0	805	8	CV120546
27	88.5	10.0	920	8	CO161709
C 28	88	9.9	551	9	DA934706
C 29	88	9.9	785	3	BQ111035
C 30	88	9.9	895	3	BQ642631
C 31	87.5	9.9	797	14	DU809059
C 32	87.5	9.9	931	3	BQ550781
C 33	87.5	9.9	1114	5	AK209417
C 34	87.5	9.9	2621	6	AK004672
35	87.5	9.9	2929	6	AK158339
36	87.5	9.9	2933	6	AK157854
C 37	87	9.8	705	2	BJ578705
C 38	87	9.8	717	2	BJ566157
C 39	87	9.8	746	13	CL393085
C 40	87	9.8	959	3	BU513771
C 41	87	9.8	988	5	CF513983
C 42	86.5	9.8	422	7	BF543156
C 43	86.5	9.8	615	2	BM265186
44	86.5	9.8	750	2	BJ974078
45	86.5	9.8	788	12	BZ578446

ALIGNMENTS

RESULT 1
CL666350/c

LOCUS
DEFINITION

CL666350
PRI0152b_A03 - PRI0152b.B21 (832) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

CL666350.1 GI:50159025

GSS.

Pristionchus pacificus

Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.

1 (bases 1 to 832)

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.

AppaDB: an AcedB database for the nematode satellite organism

Pristionchus pacificus

Nucleic Acids Res. 32 (1), D421-D422 (2004)

14681447

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: raif.sommer@tuebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end

sequenced at Vancouver, Canada.

end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1. .536
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-068J06.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"

FEATURES

source

ORIGIN

Alignment Scores:
Pred. No.: 1.92e-18 Length: 536
Score: 262.00 Matches: 51
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 29.5% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-16 (1-170) x AG192190 (1-536)

QY 6 MetSerHisProAlaLeuThrGlnLeuA:GAlaLeuArgTyrCysLysGluLeuProAla 25
|||||
Db 384 ATGTACACCCCGCGTTAACGCAACTGCGTGGCTATTGTAAGAGATCCCTGCC 443
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
|||||
Db 444 CTGGATCCGCAACTGCTGCACCTGGCTGTGGAGGATTCCATGACAAACGTTTGA 503
QY 46 GlnGlnGlyLysThrValSerValThrMetIle 56
|||||
Db 504 CAGCAGGGAACCGTAGCGTGACGATGATC 536

RESULT 4
AG192116 533 bp DNA linear GSS 06-MAR-2004
LOCUS Pan troglodytes DNA, clone: RP43-068H07.T7, genomic survey
DEFINITION
ACCESSION AG192116.1 GI:45224292
VERSION
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Pan.

REFERENCE 1
AUTHORS Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
TITLE BAC end sequences of Library RP-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 533)
AUTHORS Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);
52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.krribb.re.kr, URL:http://phs.grc.krribb.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)

Clones are derived from the chimpanzee BAC library RP-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1. .533
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-068H07.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"

ORIGIN

Alignment Scores:
Pred. No.: 1.45e-15 Length: 533
Score: 236.00 Matches: 47
Percent Similarity: 92.2% Conservative: 0
Best Local Similarity: 92.2% Mismatches: 4
Query Match: 26.6% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-16 (1-170) x AG192116 (1-533)

QY 6 MetSerHisProAlaLeuThrGlnLeuA:GAlaLeuArgTyrCysLysGluLeuProAla 25
|||||
Db 380 ATGTACACCCCGCGTTAACGCAACTGCGTGGCTCTATTGTAAGAGATCCCTGCC 439
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
|||||
Db 440 CTGGATCCGCAACTGCTGCACCTGGCTGTGGAGGATTCCATGACAAACCTTTTGGCA 499
QY 46 GlnGlnGlyLysThrValSerValThrMetIle 56
|||||
Db 500 CAGCAGGGAACCGTTATCGTGACGATGATC 532

RESULT 5

CC048367 535 bp DNA linear GSS 20-OCT-2003
LOCUS CC048367
DEFINITION O1S0415-04B1-B12 UniformMu MUTAIL Library Zea mays genomic clone
O1S-415-4-7to12-B12, genomic survey sequence.

ACCESSION CC048367
VERSION
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 515)
AUTHORS Latehaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
TITLE Sequence tagged transposon insertions from the UniformMu maize
population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110890 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:
O1S-415-4

FEATURES

source

Class: transposon insertion site.
Location/Qualifiers
1. .515
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="O1S-415-4-7to12-B12"
/clone_lib="UniformMu MUTAIL Library"

/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

ORIGIN

Alignment Scores:
 Pred. No.: 1.36e-12 Length: 515
 Score: 209.00 Matches: 40
 Percent Similarity: 97.6% Conservative: 0
 Best Local Similarity: 97.6% Mismatches: 1
 Query Match: 23.6% Indels: 0
 DB: 12 Gaps: 0

US-10-718-311-16 (1-170) x CC048367 (1-515)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
 |||||
 Db 393 ATGTCACACCCGCGTTAAACGCACTGCGTGCCTGCTATTGTACAGAGATCCCTGCC 452
 |||||
 Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
 |||||
 Db 453 CTGGATCCGCAACTGCTGACTGGCTGTGTCTGGAGGATTCATGACAAAACGTTTGA 512
 |||||
 Qy 46 Gln 46
 |||||
 Db 513 CAA 515

RESULT 6
 CL650764
 LOCUS
 DEFINITION PR10110b.P06 - PRI0110b.B21 (563) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.
 CL650764 563 bp DNA linear GSS 09-JUL-2004
 PR10110b.P06 - PRI0110b.B21 (563) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL650764
 VERSION CL650764.1 GI:50129323
 KEYWORDS GSS.

SOURCE
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 563)
 AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
 TITLE AppaDB: an AcedB database for the nematode satellite organism Pristionchus pacificus

JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)

PUBMED 14681447
 COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: raf.sommer@uebingen.mpg.de
 This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.

Seq primer: T7
 Class: fosmid ends.
 Location/Qualifiers
 1..563
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus var. California"

FEATURES
 source

Alignment Scores:
 Pred. No.: 1.46e-08 Length: 838
 Score: 171.00 Matches: 36
 Percent Similarity: 94.7% Conservative: 0
 Best Local Similarity: 94.7% Mismatches: 1
 Query Match: 19.3% Indels: 1
 DB: 13 Gaps: 0

ORIGIN

Alignment Scores:

Pred. No.: 1.18e-09 Length: 563
 Score: 183.00 Matches: 38
 Percent Similarity: 97.4% Conservative: 0
 Best Local Similarity: 97.4% Mismatches: 0
 Query Match: 20.6% Indels: 1
 DB: 13 Gaps: 0

US-10-718-311-16 (1-170) x CL650764 (1-563)

Qy 133 AspPheileGluileGlyArgAspAlaGlyLeu-TrpGlyArgSerArgLeuArgLe 152
 |||||
 Db 2 GACITTTATTGATAGCGCGTGATGCCGGCTCGTGGGGCGACGTTCCCGCTGCGATT 61
 |||||
 Qy 152 uSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 170
 |||||
 Db 62 AAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTTACCGCGCTCACCGTTGTAC 116
 |||||

RESULT 7

CL688660

LOCUS

DEFINITION CL688660 838 bp DNA linear GSS 09-JUL-2004
 PRI014a.D06.2 - PRI014a.BR (838) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL688660
 VERSION CL688660.1 GI:50197738
 KEYWORDS GSS.

SOURCE
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 838)
 AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
 TITLE AppaDB: an AcedB database for the nematode satellite organism Pristionchus pacificus

JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)

PUBMED 14681447

COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: raf.sommer@uebingen.mpg.de
 This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.

Seq primer: T7
 Class: fosmid ends.
 Location/Qualifiers
 1..838
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus var. California"

FEATURES
 source

Alignment Scores:
 Pred. No.: 4.46e-08 Length: 838
 Score: 171.00 Matches: 36
 Percent Similarity: 94.7% Conservative: 0
 Best Local Similarity: 94.7% Mismatches: 1
 Query Match: 19.3% Indels: 1
 DB: 13 Gaps: 0

ORIGIN

Alignment Scores:

US-10-718-311-16 (1-170) x CL688660 (1-838)
 Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyr-CysLysGluileProAl 25
 |||||
 Db 725 ATGTACACCCCGCGTTTAAACGCACTGCGTGCCTGCTATTTTAAAGAGATCCCTGC 784
 |||||
 Qy 25 aLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLys 42

```

Db      785  CCTGGATCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCCATGACAAA 836
|||||
RESULT 8
LOCUS   AG192305
DEFINITION Pan troglodytes DNA, clone: RP43-068M07.T7, genomic survey
sequence.
ACCESSION AG192305
VERSION   AG192305.1 GI:45224481
KEYWORDS  GSS.
SOURCE    Pan troglodytes (chimpanzee)
ORGANISM  Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE
AUTHORS  Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
TITLE    BAC end sequences of Library RP-43
JOURNAL  Unpublished
REFERENCE
AUTHORS  Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
TITLE    Direct Submission
JOURNAL  Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
Bioscience and Biotechnology (KRIIB), Genome Research Center (GRC);
52, Oun-dong, Yuseong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)
COMMENT  Clones are derived from the chimpanzee BAC library RP-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI.
R.Site 2 : EcoRI.
FEATURES
Location/Qualifiers
1..440
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-068M07.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"
ORIGIN
Alignment Scores:
Pred. No.: 2,948-07 Length: 440
Score: 160.00 Matches: 30
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 18.0% Indels: 0
DB: 14 Gaps: 0
US-10-718-311-16 (1-170) x AG192305 (1-440)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
|||||
Db 349 ATGTACACCCCGCGTTAAACGAACCTGCGTGGCTATTTGTAAGAGATCCCTGCC 408
|||||
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeu 35
|||||
Db 409 CTGGATCCGCAACTGCTCGACTGGCTGTTG 438
|||||
RESULT 9
LOCUS   BZ551713
DEFINITION pacs1-60_3287.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
pacs1-60_3287, genomic survey sequence.
ACCESSION BZ551713
VERSION   BZ551713.1 GI:27155321
KEYWORDS  GSS.
SOURCE    Pseudomonas aeruginosa
Bacteria; Gammaproteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
AUTHORS  Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE    Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL  J. Bacteriol. (2002) In press
COMMENT  Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Location/Qualifiers
1..1110
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strains="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_3287"
/clone_lib="pacs1-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."
ORIGIN
Alignment Scores:
Pred. No.: 1,428-05 Length: 1110
Score: 150.00 Matches: 45
Percent Similarity: 37.7% Conservative: 12
Best Local Similarity: 29.8% Mismatches: 56
Query Match: 16.9% Indels: 38
DB: 11 Gaps: 3
US-10-718-311-16 (1-170) x BZ551713 (1-1110)
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
|||||
Db 144 GTCGAACCCCTGCTGGAAGCTGGCAGACCCTGCCGACGAC-----GAA 188
|||||
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
|||||
Db 189 TGCCAGGGGCTCGACGTC----- 206
|||||
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
|||||
Db 207 -----CCTACCGGACGATAGCGGCTGGGTCCGCGAGGTCTAC 242
|||||
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
|||||
Db 243 CTGCATGGCCACGACCGTCCCTGGGTGTTCGCCCGGAGGTGGCGGCGCGGCGCCCTG 302
|||||
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
|||||
Db 303 GAAGGCTCGGGCTTCGACCTGGCGGTGCTGGCACCGCTCGCTGGCGAGTTGCTGTTC 362
|||||
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIle----- 137
|||||
Db 363 AGCGACAGCGCTTCGACGCGGGGCCCATCGAAGTCTGCCGCTATCCCGCGCCGCTCG 422
|||||
Qy 138 -----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGly 154
|||||
Db 423 CCCGCCGAGGTCCCGCGGAGGTCTCTGGGGCCGTCTGCTACGGTTTTCGCCGCGCGC 482
|||||
Qy 155 LysProLeuLeuLeuThrGluLeuPheLeuPro 165
|||||

```

```

Db      483 CTCGGGTGCTGGTGGCGGAGTGACTACCTACCG 515

RESULT 10
BZ564576/c
LOCUS   BZ564576             862 bp    DNA     linear     GSS 17-DEC-2002
DEFINITION pac82-164_4819.v2 pac82-164 Pseudomonas aeruginosa genomic clone
pac82-164_4819, genomic survey sequence.
ACCESSION BZ564576
VERSION   BZ564576.1  GI:27190626
SOURCE    GSS.
ORGANISM  Pseudomonas aeruginosa
          Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
          Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 862)
AUTHORS   Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
          Burns, J.L., Kaul, R. and Olsen, M.V.
          Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol. (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.
FEATURES             Location/Qualifiers
     source           1..862
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pac82-164_4819"
                     /clone_lib="pac82-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
Alignment Scores:
Pred. No.:      0.000212      Length:      862
Score:          138.00      Matches:    45
Percent Similarity: 37.3%      Conservative: 11
Best Local Similarity: 30.3%      Mismatches: 56
Query Match:    15.6%      Indels:     38
DB:             11      Gaps:        3

US-10-718-311-16 (1-170) x BZ564576 (1-862)

Qy      27 AspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGln 46
      : : : : :
Db      662 AACCCCTGCCTGGAGGCCTGGCAGACCCTGGCGGACGAC-----GAATGC 618

Qy      47 GlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIle 66
      : : : : :
Db      617 CAGGGGCTGCAGTC----- 603

Qy      67 ProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeu 86
      : : : : :
Db      602 -----CCTACCGCAGTAGCGGCTGGGTCGGCGGAGGTCTACCTG 564

Qy      87 CysAlaAspGlyGluProTrpTrpLeuAlaGlyArgThrValValProValSerThrLeuSer 106
      : : : : :
Db      563 CATGCCACGACCGCTCCCTGAGTGTTCGCGCAGCGTGGCGGCGCGAGCGCCCTGGAA 504

Qy      107 GlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThr 126
      : : : : :
Db      503 GGCTCGGGCTTCGACCTGGCGTCTCGCGACCCGCTCGCTGGGGCAGTTGCTGTTTCAGC 444

Qy      127 SerSerThrLeuThrArgAspPheIleGluIle----- 137
      : : : : :
Db      443 GACAGCGCTTCGAGCGCGGGCCCATCGAAGTCTGCGCGCTATCCGCGCGCGGCTCGGCC 384

```

```

Qy      138 -----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLys 155
      : : : : :
Db      383 GCCGAGGTCCGCGCGGAGGGTCTCTGGGGCGGCTCGCTCACGGTTTTCGCCGGCGCGCTC 324

Qy      156 ProLeuLeuLeuThrGluLeuPheLeuPro 165
      : : : : :
Db      323 GGGGTGCTGGTGGCGGAGGTGTACTACCG 294

RESULT 11
AJ863670/c
LOCUS   AJ863670             739 bp    DNA     linear     GSS 30-NOV-2005
DEFINITION Ralstonia solanacearum GSS, clone V789R, genomic survey sequence.
ACCESSION AJ863670
VERSION   AJ863670.1  GI:82937082
KEYWORDS  GSS; genome survey sequence.
SOURCE    Ralstonia solanacearum
          Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
          Burkholderiaceae; Ralstonia.
REFERENCE 1 (bases 1 to 739)
AUTHORS   Munoz-Soriano, V., Arahall, D.R., Terol, J., Buaides, C., Perez-Perez, A.,
          Liop, P., Belmonte, U.C.F., Lopez, M. and Perez-Alonso, M.
          Random genome sequencing of Ralstonia solanacearum strain IVIA 1602
          and comparative analysis with strain GMI1000
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 739)
AUTHORS   Munoz-Soriano, V.
          Direct Submission
TITLE     Submitted (18-NOV-2004) Genetica, Universidad de Valencia, C/ Dr.
          Moliner, 50, Burjassot, Valencia 46100, SPAIN
JOURNAL
FEATURES             Location/Qualifiers
     source           1..739
                     /organism="Ralstonia solanacearum"
                     /mol_type="genomic DNA"
                     /strain="IVIA 1602"
                     /db_xref="taxon:305"
                     /clone="V789R"
ORIGIN
Alignment Scores:
Pred. No.:      0.028      Length:      739
Score:          118.00      Matches:    42
Percent Similarity: 38.9%      Conservative: 23
Best Local Similarity: 25.1%      Mismatches: 64
Query Match:    13.3%      Indels:     38
DB:             14      Gaps:        3

US-10-718-311-16 (1-170) x AJ863670 (1-739)

Qy      30 LeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlyLys 49
      : : : : :
Db      599 CTGTTGATGTTGGTGAAGATCTTGAATTCGCTACGGCGCGGTCGCTGCGGCTTCCTCC 540

Qy      50 ThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGluIleProGluGlu 69
      : : : : :
Db      539 TCGTTCGCGTGAAGTCTCGTCTCAG-----CAGCGGACC 504

Qy      70 LeuProLeu-----LeuProLysGluSerArgTyrTrpLeu 81
      : : : : :
Db      503 ATGCCGTGAGCAGCAGTAGTGGCGTGGCTGGCGTTGCCCGTTCGCCGAGACGTTGGCG 444

Qy      82 ArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValPro 101
      : : : : :
Db      443 CGCAAGTCTCTGCTGATCTCGCAGCAACCCCGGTGCTTACGCACACACATCGTGCAT 384

Qy      102 ValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 121
      : : : : :
Db      383 CCGCGCAGCGTGGCGGGGACTGGCCGTTCTGAAGGGCTGGGCGACCCACCGCTCGGG 324

Qy      122 ArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGlu----- 136
      : : : : :
Db      323 CATCGCTGTTTCGCGGATCCGCGGCTTGACACGGGCGCTTCGAGTTTGGCCAGCTGGAT 264

```

```

QY 137 -----ProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGlu 91
Db 263 GTCCGGCATCGCTGTCAGCGGGCGCGTGGCGCGCTGGCGGCACGCCGCTGGCGGC 678
QY 140 AspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeu 159
Db 203 ATGGCGCGCTGCCGCGACGCCCTTCGGTATTTCGCGCGGTGCCAGCGCGATGCTGGTG 144
QY 160 ThrGluLeuPheLeuProAla 166
Db 143 ACGGAAGTGTTCGCTGCA 123

RESULT 12
LOCUS BZ578851 1141 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2 6034.xl msh Pseudomonas aeruginosa genomic clone msh2_6034,
genomic survey sequence.
ACCESSION BZ578851
VERSION BZ578851.1 GI:27213912
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1141)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Location/Qualifiers
1..1141
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2 6034"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."
ORIGIN
Alignment Scores:
Pred. No.: 1.44 Length: 1141
Score: 105.00 Matches: 44
Percent Similarity: 36.2% Conservative: 24
Best Local Similarity: 23.4% Mismatches: 70
Query Match: 11.8% Indels: 50
DB: 12 Gaps: 6
US-10-718-311-16 (1-170) x BZ578851 (1-1141)
QY 22 GluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThr 41
Db 466 CAATGTCACAGGTCCTCCCGCCACCGCTCTCGACTGGTTCGACGAGGGTCTCGTGACC 525
QY 42 LysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVal 61
Db 526 CGCGCGC-----CTGACCGTCTCGCGCACGCGCGGTTC 558
QY 62 GluGlnAsnGluIleProGluGluLeuProLeuLeu----- 73
Db 559 CGCGTGCAGACCCCTGCTGGNAGGTGGCAGACCCCTGCGCGCAGCAGCAATGCCAGGGGCTC 618

74 -----ProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGlu 91
619 GAGTCCTCCCGCGCAGTAGCGGCTGGTCCGCGAGGCTACTCTGCATGCGCCACGACCGT 678
92 ProTrpLeuAlaGlyArgThrValValProValSerThrLeuSerGlyPro----- 108
679 CCTGGGTGNTCCCGCGCAGCGTGGCGCGCGCGCTGAAGGCTCGGCTTCGACT 738
109 -----GluLeuAlaLeuGlnLysLeuGlyLysThrPro 119
739 TGGCGCTGCTCGACCCGCTGTTGGCGAGTTCTTGTAGGAAGCGCCTTGAGCGCGGC 798
120 LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIle-----GluIle 137
799 CCATCAAGACTGGCGTATCCGCGCGCGCTTTCGCGCGCAGTTCGCCGGGAGGCTCTCTG 858
138 GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArg-----Leu 152
859 GGCCGCTCTTACGGTTTCCCGCGCGCGCTGGGTGCTGTGGCGAAGGTACTACCGGCGCTG 918
153 SerGlyLysProLeuLeuLeuThr----- 160
919 GGGACAGCGCCGGAATTGCGGACGATAAACCCCGCGCTGAATTCGGAAGTCCCAAGTTTGG 978
161 GluLeuPheLeuProAlaSerPro 168
979 AACCTTATCAAGCCCTTGTCCCT 1002

RESULT 13
LOCUS CL696699/753 753 bp DNA linear GSS 10-JUL-2004
DEFINITION PRI019d_B03_2 - PRI019d_BR (753) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL696699
VERSION CL696699.1 GI:50218607
KEYWORDS Pristionchus pacificus
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 753)
AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
TITLE AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
PUBMED 14681447
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
FEATURES
Location/Qualifiers
1..753
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"
ORIGIN
Alignment Scores:
Pred. No.: 10.2 Length: 753
Score: 95.00 Matches: 19
Percent Similarity: 100.0% Conservative: 0

```

```

Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      10.7%      Indels: 0
DB:              13          Gaps: 0

US-10-718-311-16 (1-170) x CL696699 (1-753)

QY 152 LeuSerGlyLeuProLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 170
DB 751 TTAAGCGGTAAACCGCTTGTCTACAGAACTGTTTTTACCAGCGTCACCGTTGTAC 695

RESULT 14
LOCUS BG799182 591 bp mRNA linear EST 21-MAY-2001
DEFINITION f3jic10.y1 zebrafish gridded kidney Danio rerio cDNA clone
IMAGE:4744722.5 similar to TR:Q9Y4G8 Q9Y4G8 KIAA0313 PROTEIN. ;,
mRNA sequence.
ACCESSION BG799182
VERSION BG799182.1 GI:14163514
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 581)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
WASHU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
Washington University Genome Sequencing Center Clone distribution:
Genome Systems, St. Louis, Missouri (web address:
www.genomesystems.com) (email contact: info@genomesystems.com) and
Research Genetics, Huntsville, Alabama (web address:
www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
High quality sequence stop: 473.
FEATURES
Location/Qualifiers
source
1..581
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:4744722"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XL0LR"
/clone_lib="zebrafish gridded kidney"
/notes="Organ: kidney; Vector: pBK-CMV; Site 1: EcoRI;
Site 2: XhoI; Oligo dt cDNA library constructed from mRNA
pooled from pooled kidney tissue from 300 adult
zebrafish."

ORIGIN
Alignment Scores:
Pred. No.: 11.7 Length: 581
Score: 93.00 Matches: 38
Percent Similarity: 45.4% Conservative: 26
Best Local Similarity: 27.0% Mismatches: 43
Query Match: 10.5% Indels: 34
DB: 2 Gaps: 7

US-10-718-311-16 (1-170) x BG799182 (1-581)

QY 7 SerHisProAlaLeuThr-----GlnLeuArgAlaLeuArgTyrCysGlyeGluLeuPro 24
DB 117 AGTAACCCCTGACCTGGCACAGGCGCAGCAAGCATCATCGACTACAGTACACAGCCACCA 176

QY 25 AlaLeuAspProGlnLeuLeu-----AspTrpLeuLeu 35
DB 177 GAGTTGCAAGACCAGGTGCTCGCATTTTCAAGCAGATCAGCAAGCCGCTTACTTACTGT 236

QY 36 LeuGluAspSerMetThrLysArg-----PheGluGlnGlnGlyLys 49
DB 237 GCCAATAGACACACACAGCCAGAGAGCTGCCAATCTAGCCATAAAGAGTTTGCCTTG 296

QY 50 ThrValSer-----ValThrMetIleArgGluGlyPheVal 61
DB 297 TCTCGAGTCCCGAAGCTTTTTCGCTCTGTGAAGTTTCAGTCACACAGGAAGGAGTTATC 356

QY 62 GluGlnAsnGluIleProGluGluLeuProLeuLeuProLys-----GluSer 77
DB 357 AAGCAGAGACGACTGCCTGATCAACTGTCAACTGGCTGACAGATCCAACCTGAGTGCC 416

QY 78 ArgTyrTrpLeuArgGluIleLeu-----LeuCysAlaAspGlyGluProTrp 93
DB 417 AGGTACTATCTAAAGACCAATATGGAACCGAGACGCTGTCTCGGATGTGGAAGCACTG 476

QY 94 ---LeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeu 112
DB 477 GAGATGCAGAGAGAATCTGTGTCGCCCTGTGTCTTTGAGCTCAATGGAGATAGCCAAT 536

QY 113 Gln 113
DB 537 CAG 539

RESULT 15
LOCUS CR861280 4527 bp mRNA linear HTC 12-NOV-2004
DEFINITION Pongo pygmaeus mRNA; cDNA DKFZp459N063 (from clone DKFZp459N063).
ACCESSION CR861280
VERSION CR861280.1 GI:55733331
KEYWORDS HTC.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Pongo.
REFERENCE 1 (bases 1 to 4527)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Oeanger,A.,
Fobo,G., Han,M. and Wiemann,S.
CONSRM The German CDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Agowa (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp459N063) is available at the RZPD Deutsches
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp459N063
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.
FEATURES
Location/Qualifiers
source
1..4527
/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp459N063"
/tissue_type="cortex"
/clone_lib="459 (synonym: pcor1). Vector pSport1_Sfi; host
DH10B; sites SfilA + SfilB"
/dev_stage="adult"
/notes="ATPase, H+ transporting, lysosomal accessory

```

Gene protein precursor (Homo sapiens) "

1..4527

/gene="DKFZp459N063"

17..367

/gene="DKFZp459N063"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAH93348.1"

/db_xref="GI:55733332"

/translation="MMAATATATVRMGPRCAQALWEMPMPLPVFLSLAAAVVAARAAA

VAEQVPLVLMSSDRDLNAPADTHEGHITSDQLSTYLDPALELGPRLVLLFLQDKV

RPFPFLPQSSGS"

ORIGIN

Alignment Scores:

Pred. No.: 252 Length: 4527
Score: 92.50 Matches: 48
Percent Similarity: 35.1% Conservative: 23
Best Local Similarity: 23.8% Mismatches: 76
Query Match: 10.4% Indels: 55
DB: 6 Gaps: 8

US-10-718-311-16 (1-170) x CR861280 (1-4527)

QY 4 TrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeu----- 17
Db TGGAGGGCAAGTTCGGCAGCTGCTACAAAAAGCCAGTATCACCTGTGATCCTCTCC 3270
QY 18 -----ArgTyrCysLysGluIleProAlaLeuAspProGlnLeu----- 31
Db TGTGAGTTACACGACACCGCTCCCGGATCTGTTCGGGCCAAACTTCTCTGTGGC 3330
QY 31 ----- 31
Db GTACAGGACCAAGTGGAGGAGCTGACTCCCTCACCCTTTGGGTGCAGAGCTCAACCT 3390
QY 32 AspTrpLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 51
Db GACTGGCTCTTCTGGATGACTCCTTTGCCAGGCTCTCACTGACCTATGAACGACTCTT 3450
QY 52 SerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeuPro 71
Db TGGTACCACAGTGACATTCAGGTCTATCTGGCCACACCGCCTCTACCCAGTATCTGCCG 3510
QY 72 LeuLeu-----ProLysGluSerArgTyrTrpLeuArgGluIleLeuLeu 86
Db GCACGTGGTTTACATGGAGCGCTCGAAGTCCACAGCAATGGCTCGTCGCTACTTCAA 3570
QY 87 CysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSer 106
Db TGCTTCCCAAGGTAC-----AGGGCCACAGCATCTACTCTTCCACTGGAGTATGTCTAG 3624
QY 107 GlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThr 126
Db CAGCCTGAG-----CAAGAAGGGGAGTCTCTCTGCGCCCGCAC----- 3663
QY 127 SerSerThrThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyArg 146
Db GCAGCCCTCTCTCTGGCAGATATGCTTCAGGACTTCCAGATCCAGGCTTT----- 3714
QY 147 ArgSerArgLeuArgLeuSerGlyLysProLeuLeu-----LeuThrGluLeuPhe 163
Db -----CAACGTAATGGGGAGAGAGTTCTTCTACGCCAGCGACTGTGCCAGCTTCTT 3765
QY 164 LeuPro 165
Db CTCCCC 3771

Search completed: May 30, 2006, 04:08:17

Job time : 4396.04 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)